Estimation of Quality and Intelligibility of a Speech Signal with varying forms of Additive Noise

[1]V Venkata Narayana [2]Sk Hasane Ahammad [3]B. Vinay Chandu, [4]G.Rupesh, [5]G.Abishek Naidu, [6]G.Pavan Gopal

[1,2,3,4,5,6] Koneru Lakshmaiah Education Foundation, Vaddeswaram, India, ahammadklu@gmail.com

ABSTRACT

 In this concerned article, we will focus on performance and comprehensibility improvement strategies. To seek the masking skills, different types of techniques are used. Masking is used to reduce noise in the voice signal and specific sounds are applied to the voice signal. Different techniques were developed to compensate different noises, whether stationary or non-stationary. It provides two types of algorithms in general. Periodic signal and combination filter are the first algorithm. There, a regular signal combines the speech signal, and the output is sent through the filter of the mixture. The ends of any audio signal are used for the second type of algorithm. The second algorithm uses some noise (white(or) brown(or) pink) modulated by the voice envelope on the amplitude.

Key words : Combinational filter, extremity, intelligibility, masking, non-stationary noise, stationary noise, periodic signal, speech envelope**.**

1. INTRODUCTION

The speech signal's accuracy depends primarily on its smartness. Intelligibility is a outcome of a listener's listening quality of voice. Any distortions of speech caused by noise will adversely affect listeners ' perception of speech [1-3]. Which means it has an impact on listeners ' intelligibility. Research, for instance, show that noise disruption in the form of closely seated student discussions is far more harmful than reverberation action. Sound masking (or) speech masking blends natural or artificial sounds which are commonly known by the names of "white noise" or "pink noise", to conceal unwanted sounds in a room. This sound masking process reduces awareness in a particular area of existing sounds. Such shades of noise (brown, red, white) refers to a noise signal's power spectrum. Similar noise colours have different characteristics. Audio signals will sound different from human ears, and the structure of objects will be distinctly different [6]. Usually, each application requires a specific colour noise. The naming of noise types after colours, a symbol whose amplitude is equal to energy at any frequency interval, has its origin with white noise. The naming was determined by a contrast between white light and the visible range with such a flat spectrum. Certain names for colours, such as purple, green, were then given to noise with other spectral profiles, often referring to the colour of light with different spectra The White noise is known as any type of audio signal given a constant spectral density of energy at different frequencies of equal frequency [8-10]. Inside of the frequency spectrum, pink noise is known as a signal, for which, if it's power levels of each frequency interval is increased, the signal frequency of it will show a decline and if the signal frequency is increased, the power levels will show a decline [8-10]. When many sound particles move at a different pace and also in random directions, we call that type of motion which is encountered by them as Brownian motion. Brown (or) red noise follows in steps of Brownian motion, so, its alternatively called as random walk noise [8-10]. Although the same signal is represented by these sounds, they have their own variations.

The noise disparity is illustrated are White Noise is widely distributed throughout the sound spectrum, including sounds of low frequency, medium, high frequency. Example: frequency ranges between 39 Hz-60 Hz contains same amount of sound power as the sound power range between 390 Hz and 410 Hz, as both intervals are 20 Hz wide [9-14].

Considering the low frequency extremity of the spectrum, pink noise becomes louder and quieter at the high frequency extremity. Example: Considering the frequency ranges in between 50 to 70 Hz, pink noise would have the same power as in the band from 3000 to 5000 Hz.

At the end of the lower spectrum, the brown (or) red noise is very powerful and deep regardless of white, pink noises. There are definitely two contrasting types of protruding noise, which are stationary noise and non-stationary noise. Any noise known as stationary noise has the probability that the noise voltage lies within any given interval is constant. Random noise defined as non-stationary noise has the probability that the noise voltage is not within the same interval and that always varies with time.

2. THEORETICAL PROCEDURE

The speech signal will be taken as a sample of any nature and will be checked against the method of masking. For the existence of any noise disturbance, the speech signal SNR will be measured.

The forms of noise which are depicted as black and brown and also pink noises, were used to minimize the interference in order to find out the correct masking technique.

A Stationary Noise Masking Ability Measure

 Stationary discreet white, purple, and brown noise were created by MATLAB. Coloured (pink and brown) noise

composed of 7 octave band filters with core frequencies 125, 200, 450, 1500, 2500, 4500 and 7500 Hz, moving white noise through a filter stack. Sounds which are generated through audio signals have their power spectrums in the shape of step. The desired phase heights were regulated in each frequency channel through the selection of gain values. Clear speech x(t) was applied to the optimal SNR n(t)

$y(t) = x(t) + k_1 \times n(t)$

Here, $k = 10^{0.05(SNR_1 - SNR_0)}$

k=weight coefficient [4], SNR_1 = signal-noise ratio for normalized noise.

White noise has the worst, relative to pink and brown noise, at poor signal-to-noise ratios (less than 0 dB). The pink noise masking skill is the best [7]. Of SNR>5dB, the brown noise is slightly better than white noise.

B. Nonstationary Noise Masking Abilities

HSLN (Human-speech as noise) can be generated, uncorrelated with the masking signal, from both the masking signal and the speech fragments. In the concerning article, we are confined to focus on the noise generated by the speech of masking.

$$N_1(t) = [x(t) \times m(t)] Ä h(t)$$

where,

$x(t)$= speech signal
$m(t)$= given shape periodic carrier
$h(t)$=comb filter impulse response
$\square$ =convolution symbol

Here the material x(t).m(t) defines the balanced modulation procedure that provides a complex voice signal spectrum inversion. Choosing the carrier m(t)'s T0 cycle allows you to handle a signal spectrum f(X) shift value along the frequency axis. Selection of carrier form m(t) allows one control weight to overlay the shifted spectrum f(x) repeatedly. The second algorithm's important feature is that the, signal-noise ratio is virtually constant over time. As a result, during speech breaks, noise is almost not detected, allowing sound pollution to be reduced.

3. EXPERIMENTAL RESULTS
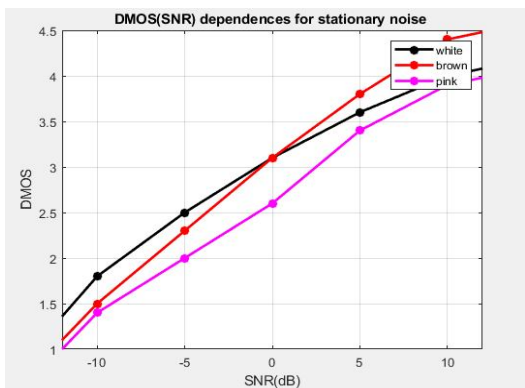
A. Stationary Noise

Figure. 1: Depiction of each signal in amplitude on behalf of time.

In the above figure 1, we're considering a speech signal and mixing it with the noise signal. After mixing the audio form of speech signal with the noise signal, noisy input signal is taken as the updated input. Finally, an estimated output signal which is free of noise is depicted. Each and every signal is depicted on behalf of its amplitude with respect to time and are denoted by the formation of a graph.
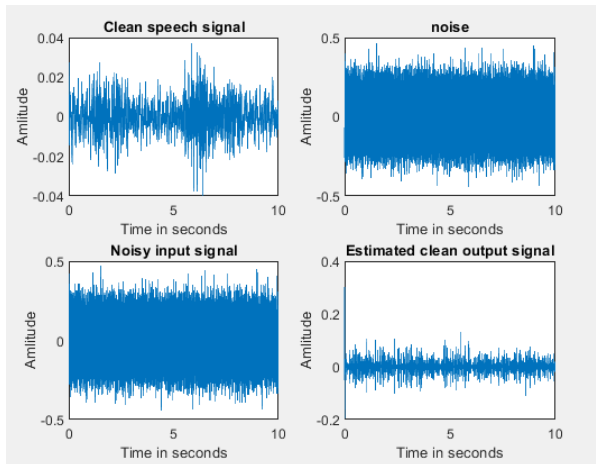

Figure. 2: Depiction of Dependencies of Stationary noise.

Figure 2 shows the results of performance assessment of speech signals distorted by stationary noise. From Fig as can be seen. 2. White noise has the worst, relative to pink and brown noise, masking at poor signal-noise ratios (less than 0 dB). The brown sound is marginally better than the| white sound of snr>5db.

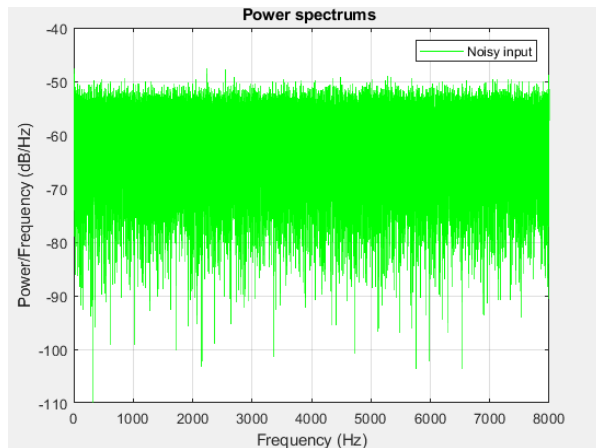## B. Nonstationary Noise

Figure. 3: Depiction of PSD of input signal.

In the figure 3 above, the output signal is a clear signal of speech.
This signal can be in any form which includes noise. Now, in the diagram above, power levels of input speech signal is shown.
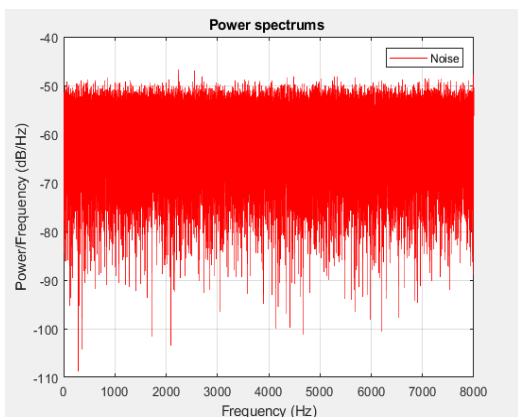
Figure. 4: Depiction of PSD of noise signal

In the above figure. 4, the input signal is taken in the form of a noise signal. This signal can be in any form of noise. Now, the PSD [5] of the input noise signal is depicted in the above graph.
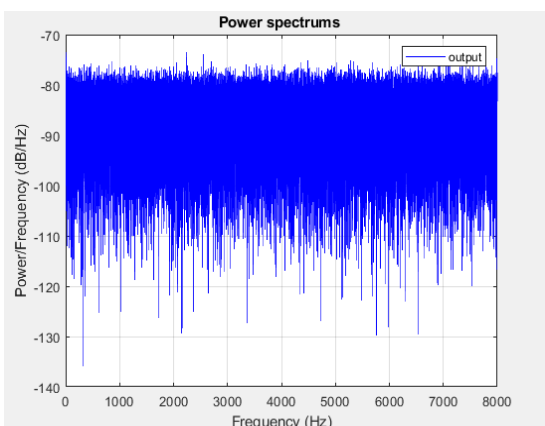

Figure. 5: Depiction of PSD of noisy input signal

In the above figure. 5, the input signal is taken in the combination of both clear speech signal and noise signal. After the addition of both the signals, PSD[5] can be found out. The spectral power of the noisy input signal is now shown by forming graph such as above.
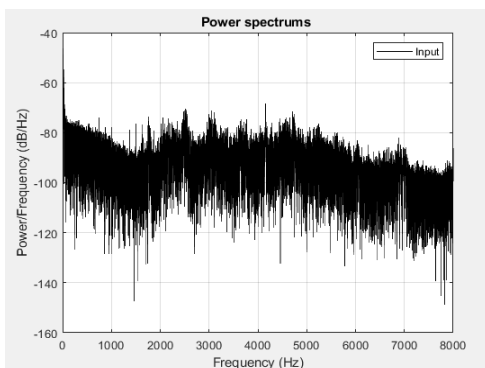

Figure. 6: Depiction of PSD of output signal

After the removal of presence of noise, the signal which we get after the process is coined as the output signal. In the

above figure. 6, the output signal is taken in for finding out the PSD. Now, the PSD [5] of the output signal is depicted in the above figure 6.
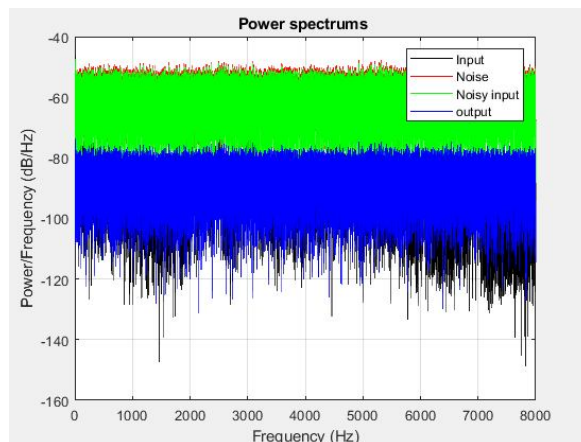

Figure. 7: Combined Depiction of PSD of all signals

The combined depiction of the PSD's of the signals is depicted above. As you can see, the improvement of the PSD[5] of these signals is very variable
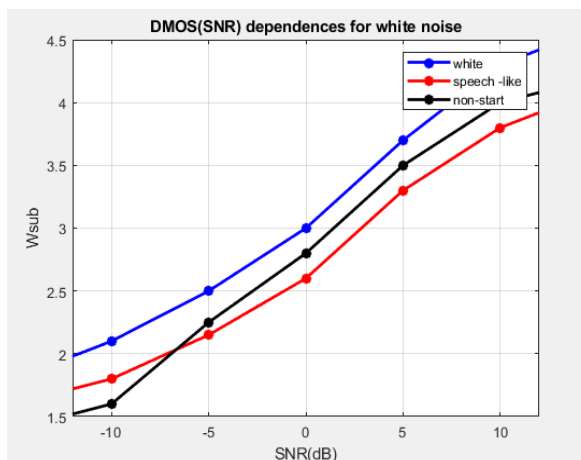

Figure. 8: Depiction of dependencies of white noise

As you can see from Figure 8, proposed in this paper, noise despite the relatively simple realization of their algorithms, have better masking properties in the range SNR=-16... +16dB compared to white noise. The fact that the SNR is almost constant in time for considered noise may explain this result. In other words, if the intensity of a audio signal increases, at that instant, the intensity of noise decreases.

4. CONCLUSION

Several noise models ' effect on speech performance and their intelligibility were observed and learned. It is shown that the results of concerning subjective speech quality assessment are in same boat with those of objective evaluation of speech

intelligibility for white, and pink and brown noise. This shows that it is important to use simple, relatively inexpensive methods to evaluate speech content against the high usage of labour and very expensive approach of evaluating subjective speech intelligibility. Masking capabilities of two simple non-stationary noise types were studied for generation. Next, non-stationary noise masking function is lower than one for white noise. Second, during speech breaks, the air is not contaminated by excessive noise. Discrete white noise which is depicted by w(t) was used as carrier as the simplest to implement.

REFERENCES

1 J. R. Reich, Bradley, and S. Norcross, **Speech Intelligibility test on the mixed Effects of Signal-Noise Ratio and Room sounding system**, J. Acoust. It's Soc. Am, Vol. 106, 1999.
https://doi.org/10.1121/1.427932

2 H. Sato, J. Bradley, **Acoustic Condition Assessing test for Speech(audio)Communication in the Classrooms in a Working Elementary School**, J. Acoust-Soc. Am., 4, 106, 2004.
https://doi.org/10.1121/1.4786138

3 J. Bradley, H. Sato, **Experimental Speech quality Test Results for Classrooms of 1st and 3rd and 6th Grade Boys**, ICA proceedings, Kyoto, 2004.

4 weighted signal to noise ratio [Online]. Available: https://dsp.stackexchange.com/q/44615 power spectral density mat lab [Online].

5 A. Horev, Y. Makarov, "**On Acoustic Data Leak Prevention Effectiveness**," Specific Technology, No. 5, pp. 46-56, 2000. (in Russian).

6 Ahammad, S. H., V. Rajesh, K. Saikumar, S. Jalakam, and G. N. S. Kumar. 2019. **Statistical Analysis of Spinal Cord Injury Severity Detection on High Dimensional MRI Data**. International Journal of Electrical and Computer Engineering 9 (5): 3457-3464.

7 Speech technology for telecommunications F.A. Westall , R. Denis Johnston , Alwyn V. Lewis Sound Masking Jesse Russell , Ronald Cohn **, An Experiential Approach to Sound, Music, and Psychoacoustics** Why You Hear what You Hear: Eric J. Heller

8 M. Surhone, Miriam T. Timpledon, Susan F. Marseken. **Sound Masking: White Noise, Pink Noise, Noise Pollution, Auditory Masking, Active Noise Control, Architectural Acoustics, Health Effects** From Noise, Noise Mitigation, Noise Reduction Coefficient Lambert

9 Hasane Ahammad, S. K., & Rajesh, V. (2018). **Image processing based segmentation techniques for spinal cord in MRI.** Indian Journal of Public Health Research and Development, 9(6), 317-323.
https://doi.org/10.5958/0976-5506.2018.00571.5

10 Ahammad, S. K. H., Rajesh, V., & Ur Rahman, M. Z. (2019). **Fast and accurate feature extraction-based segmentation framework for spinal cord injury severity classification**. IEEE Access, 7, 46092-46103. doi:10.1109/ACCESS.2019.2909583

11 Vijaykumar, G., Gantala, A., Gade, M. S. L., Anjaneyulu, P., & Ahammad, S. H. (2017). **Microcontroller based heartbeat monitoring and display on PC. Journal of Advanced Research in Dynamical and Control Systems,** 9(4), 250-260.

12 Raj Kumar, A., Kumar, G. N. S., Chithanoori, J. K., Mallik, K. S. K., Srinivas, P., & Hasane Ahammad, S. (2019). **Design and analysis of a heavy vehicle chassis by using E-glass epoxy & S-2 glass materials. International Journal of Recent Technology and Engineering, 7(6), 903-905.**

13 Hasane Ahammad, S., Rajesh, V., Hanumatsai, N., Venumadhav, A., Sasank, N. S. S., Bhargav Gupta, K. K., & Inithiyaz, S. (2019). **MRI image training and finding acute spine injury with the help of hemorrhagic and non hemorrhagic rope wounds method**. Indian Journal of Public Health Research and Development, 10(7), 404-408.
https://doi.org/10.5958/0976-5506.2019.01603.6

14 Saikumar, K., Rajesh, V., Ramya, N., Ahammad, S. H., & Kumar, G. N. S. (2019). **A deep learning process for spine and heart segmentation using pixel-based convolutional networks**. Journal of International Pharmaceutical Research, 46, 278-282.