AN EFFICIENT MODIFIED CUCKOO SEARCH BASED NOISE DESTRUCTION AND IMPROVEMENT OF SPEECH SIGNAL QUALITY

Anoop V[1], PV Rao[2]

[1]Research Scholar, Dept. of ECE T John Institute of Technology, Bangalore, Karnataka-India
[2]Professor and R&D Head, Dept. of ECE, Rajarajeswari Engineering College, Bangalore, Karnataka, India

Abstract

The harmonic signal takes its origin from a restricted spectral illustration of a voice signal. It is integrated with at least a section of the restricted delayed spectral signal to furnish a reconstructed speech signal equipped with conceptually enhanced audio quality. Speech signal quality augmentation involves the challenge to scale down the noise and to perk up the transparency. In this document, we valiantly launch an effective noise destruction method for perfection of speech signals by means of the modified cuckoo search algorithm. At first stage, the Input original signal and noisy speech signal features are extracted using the Amplitude Magnitude Spectrogram (AMS). Then the signals are classified in order to generate the initial population. The Second stage is to generate the best weight using modified cuckoo search algorithm and in the third stage, Ranking and Merging is used in order to improve the quality by ranking the best solution from the merging of Modified levy flight and Velocity in PSO. Hence the underlying objective, in this regard, is to boost the speech signal quality and do away with the noise prevalent in the signal. The cheering outcomes achieved have unequivocally ratified effectiveness of our proposed technique and is able to destruct noise and to improve the quality of the speech signal

Keywords: Amplitude Magnitude Spectrogram, Modified Cuckoo Search, PSO, and SNR

--***--

1. INTRODUCTION

The overall quality, in effect, represents the overall impression of the listener on how "good" the quality of the speech is. And it is the listener who has the power to define the term "good". Nevertheless, as we are accustomed to hearing genuine air-transmitted speech originating from various people on a daily basis, this speech furnishes a "reference point" on the quality range. The listeners evaluate the speech under investigation based on this reference. Speech enhancement (SE) is intended to augment the efficiency of speech communication systems in noise-polluted circumstances [8]. Let us take for instance, SE executed in a mobile radio communication system, a speech to text system, a speech detection mechanism, a set of low quality recordings, or a system to enhance the functioning of aids for the hearing impaired [9]. In fact, SE has surfaced as a standard issue in signal processing. Hearing aid users habitually experience great intricacy in comprehending speech in an earsplitting backdrop. They characteristically need a signal-to-noise ratio SNR of about 5–10 dB which is superior to that of standard hearing listeners to attain an identical grade of speech comprehension [11]. No wonder, various solitary and multi-microphone noise decrease techniques have been devised and launched for modern hearing aids. Multi-microphone noise decline mechanisms are competent to utilize spatial along with spectral data and are hence are favored to single-microphone systems [12].

The methodology used in this paper for noise destruction and speech signal quality improvement system consists of three stages namely; Feature extraction stage, Modified CS based generating finest weight stage, Merging and ranking stage. Initially, the original and noise speech signals are given as input to extract features and by using the modified cuckoo search algorithm finest weight can be evaluated, merging and ranking unit is used to get the improved quality signal. The rest of the paper is arranged as follows a short description of modified Cuckoo search algorithm is illustrated in Section 2. The test outcomes and relative analysis debate are offered in Section 3. At last, the conclusions are furnished in Section 4.

2. SPEECH SIGNAL QUALITY IMPROVEMENT USING MODIFIED CUCKOO SEARCH ALGORITHM

The methodology used in this paper for noise suppression and speech enhancement system consists of three stages namely; Feature extraction stage, Modified CS based generating finest weight stage, Merging and ranking stage. Initially, the original and noise speech signal is given as input to extract features and by using the modified cuckoo search algorithm finest weight can be computed and merging and ranking unit is used to get the improved quality signal

2.1 Cuckoo Search Algorithm

Cuckoo birds invite attention in view of their distinctive aggressive reproduction technique. Cuckoos tend to employ brood parasitism, in which a bird (brood parasite) lays and leaves its fundamental kinds of brood parasitism such as intra-specific brood parasitism, cooperative breeding, and nest takeover. Certain species like the Ani and Guira cuckoos lay their eggs in communal nests, even though they are likely to take away others' eggs to enhance the hatching possibility of their own eggs. Certain host birds are not friendly vis-a-vis intruders and indulge in straight fight with them. In such cases, host bird discards the unfamiliar eggs. In certain other occasions, further sociable hosts just leave its nest and create a new nest somewhere else. Cuckoo Search algorithm is one of the outstanding population based stochastic global search meta-heuristics method. In this algorithm, potential solutions represent Cuckoo eggs. Natural systems are complicated and hence, cannot be designed by computer algorithms in its fundamental form. Generalization of natural systems is highly essential for successful execution of computer algorithms.

2.2 Modified Cuckoo Search Algorithm

Modified Cuckoo search (CS) is one of the modern optimization algorithms. For a maximization issue, the quality or fitness of a solution is basically proportional to the value of the objective function. In the bona fide world, if a cuckoo's egg is very similar to a host's eggs, then this cuckoo's egg has minimal possibility of being found out, thus the fitness has to be linked to the variance in solutions. Hence, it is advisable to perform a random walk in a prejudiced manner with certain random step sizes. In fact original and modified code employs arbitrary step dimensions. In comparison with the original code, we employ diverse function sets for computing the related step dimension. The algorithm is rooted on the tendency of certain cuckoo species in conjunction with the modified Levy flight behavior of some birds and fruit flies and Velocity from PSO for added augmentation of the best solution.

2.3 Feature Extraction Stage

In feature extraction stage, features are extracted from the input speech by means of the Amplitude Magnitude Spectrogram (AMS). The input speech signal is a blend of Original speech signal and the loud signal. The input signal is at first processed by carrying out three estimations so as to adapt the signal fit for advanced processing.

- ❖ Sampling
- ❖ Quantization
- ❖ Pre-emphasized

Each and every channel is characterized by the upper limit frequency U_i and the lower limit frequency L_i. After building the channel bands, envelope of each band is computed by the full wave rectification and thereafter, the envelope is decimated by a factor of 3 which is subsequently fragmented into overlapping fractions of 128 samples of 32 ms with an overlap of 64 samples. When the fragmentation is completed the sampled signals achieved are windowed by means of Hanning with a view do away with unnecessary signal segments and to attain sharper summits

$$Y_k = \sum_{k=0}^{N-1} y_n e^{-j2\pi k \frac{n}{N}}, where \, k = 0,1....N-1 \quad (1)$$

Where, N characterizes the width of the samples in the symmetrical window function

n -Integer, with values changing from $0 < n < N-1$.

The time delta functions ΔA_T as furnished below.

$$\Delta G_T(\lambda,\phi) = G_F(\lambda,\phi) - G_F(\lambda,\phi-1), where \, \phi = 2,...,T \quad (2)$$

The overall feature vector $G(\lambda,\phi)$ including the delta functions may be defined as

$$G(\lambda,\phi) = [G_F(\lambda,\phi), \Delta G_T(\lambda,\phi), \Delta G_S(\lambda,\phi)] \quad (3)$$

Therefore, we have extracted the features from a mammoth speech signal corpus by means of AMS feature extraction. Thus the intention of producing best weight is to scale up the speech signal quality. Now, the input stage is classified into the respective classes by using the original signal and noisy signal. The categorization of the speech signal in to diverse classes is in accordance with the Quality Ratio which is the ratio of the estimated speech magnitude $\overline{M}$ to the true speech magnitude T for each T-F unit. Here, the spectrum at time slot ϕ and sub-band λ are taken into account and therefore the quality ratio R_Q can be expressed as

$$R_Q = \frac{|\overline{M}(\lambda,\phi)|}{|T(\lambda,\phi)|} \quad (4)$$

Where estimated signal spectrum $\overline{M}$ is achieved by the product of spectrum M with the gain function G_A which is shown in the equation below:

$$\overline{M}(\lambda,\phi) = G(\lambda,\phi).|M(\lambda,\phi)| \quad (5)$$

Where Gain may be estimated from the equation:

$$G_A(\lambda,\phi) = \sqrt{\frac{\psi(\lambda,\phi)}{1+\psi(\lambda,\phi)}} \quad (6)$$

2.4 Modified Cuckoo Search for Generating the Finest Weight Stage

2.4.1 Initialization

At the initial noisy speech Input signal is represented by X, which is defined by $X = \{x_1, x_2, ..., x_{Ys}\}$

Where, Ys - is the total number of signals in the input signal The input signal is classified into stages such as P_1, P_2 or P_3 by employing the quality ratio. Moreover, so as to locate a finest weight with minimal number of iterations, at the outset we must classify the components into several stages and thereafter attain the initial solution with the aid of categorization segment. Subsequently, the Signal to Noise ratio (fitness) is estimated for the preliminary population.

2.4.2 Obtaining New Solutions Using Modified Levy Flights:

The modified cuckoo search is executed with an eye on attaining a new solution by means of modified Levy flight equation and Velocity update in PSO

Let us consider the signal M_{i1} in M_i, and then the enhanced value representing the new solution m_{i1}^{*} is furnished by

The Levy flights used for ordinary cuckoo search algorithm which is shown below:

$$m_{i1}^{*} = m_{i1}^{(t+1)} = m_{i1}^{(t)} + \alpha \oplus Levy(n) \qquad (7)$$

By altering the above equation i.e. Levy flight equation by employing the gauss distribution we get

$$m_{i1}^{*} = m_{i1}^{(t+1)} = m_{i1}^{(t)} + \alpha \oplus \sigma_s \qquad (8)$$

Where, $\sigma_s = \sigma_0 \exp(-\mu\kappa) \qquad (9)$

Where,

σ_0, μ - Constants

κ - Current generation

$\alpha > 0$ - is the step dimension which should be linked to the issue of concern. It can be employed in several cases, with the intention of accommodating the difference between solution qualities. The changes to the technique are carried out with the objective of enhancing the convergence rate and accuracy, though in the process the smart traits of the original technique are lost. Therefore, we resort to getting new solutions and thereafter the fitness value (SNR value) of the new solution is estimated.

2.4.3 Obtaining New Solutions Using Velocity of PSO:

In each and every generation, the preliminary velocities are updated and based on the updated velocity, the particles are also updated. The velocities and particles updated are shown below:

$$V_i^{(it+1)} = V_i^{(it)} + l_1 * r_1(pb_i - c_i^{(it)}) + l_2 * r_2(gb_i - c_i^{(it)}) \qquad (10)$$

$$c_i^{(it+1)} = c_i^{(it)} + V_i^{(it+1)} \qquad (11)$$

The particles update their location and the velocity until it attains its stop condition. The procedure is replicated till the highest number of iterations is realized. When the greatest number of iterations is attained the procedure gets instantly finished.

3.5 Merging And Ranking The Best Solution Stage:

Immediately when the estimation and substitutions are made, we turn to do away with a portion of worst nests and produce fresh nests in their place, which is performed locating the quality of the entire current nests and evaluating it, by way of preserving the best solutions and substituting the worst nests by freshly constructed nests. Accordingly the solutions are graded and the current best is estimated. The entire loop goes on till certain stop standard is satisfied and the current best in the previous loop executed becomes the best solution. By merging the best values from modified levy flights and velocity in PSO, we have graded the best fitness solution as the finest one. Therefore, the resultant signals are fused to formulate the improved speech waveforms. Now, at the outset, the noisy speech signal is multiplied with the finest weight produced from cuckoo search algorithm directly.

$$E_s(k,t) = B(k,t) * N(k,t) \qquad (12)$$

Where,

$N_s(k,t)$ -Noisy speech signal

$B(k,t)$ -Best weight generated

$E_s(k,t)$ - Enhanced signal

Therefore, at last, the original speech signal is evaluating after adding the weighted responses of the 25 signal segments. The spectrogram of the estimated speech signal illustrates the level of energy akin to the original speech signal energy level at the related frequencies.

4. RESULTS AND DISCUSSION

The outcomes attained and the resulting debates based on them are detailed in this section. The investigational set up

and replication outcomes are thrashed out below. The database is used for obtaining noise sounds including suburban train noise, babble, car, exhibition hall, restaurant, street, and airport and train-station noise. The sentences were initially sampled at 25 kHz and then down sampled to 8 kHz and noise segment is synthetically supplemented to the speech signal.

4.1 Experimental Set Up and Simulation Results

The proposed technique for speech enhancement is implemented in a system having 8 GB RAM with 32 bit operating system having i5 Processor using MATLAB Version 2011. The signal power is plotted for a frequency range between 0 to 2.5 KHz. The input signal, noisy signal and the de-noised signal are shown in below figure.
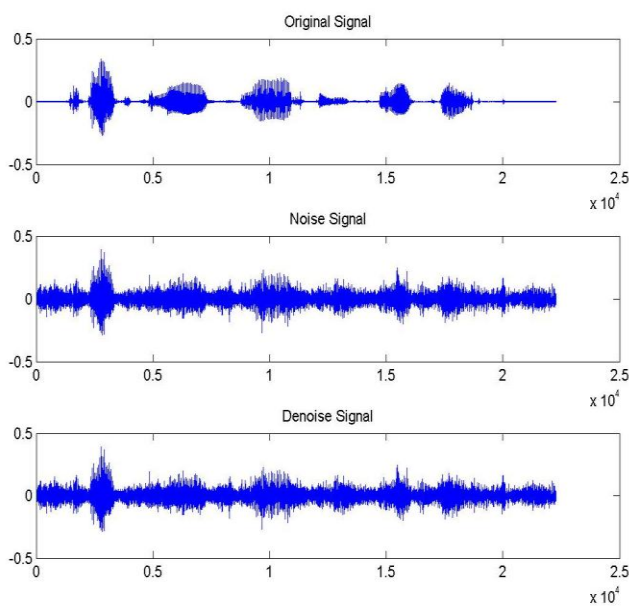
Fig 1: Input signal, Noise signal and Denoised signal for 0 dB babble Noise
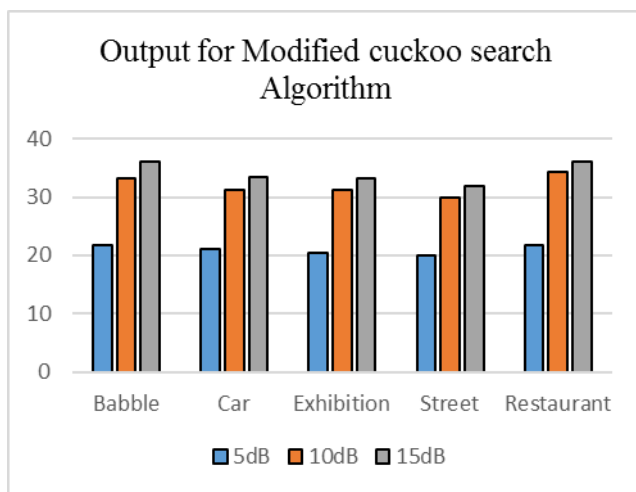
Fig 2: SNR values for Proposed Babble, Car, Exhibition, Street, and Restaurant noises of 5db, 10dB and 15dB using Modified Cuckoo Search

Table 1: Comparison for SNR values of Proposed and Existing method for the Babble, Car, Exhibition, Street, and Restaurant noises of 5db, 10dB and 15dB using Modified Cuckoo Search Algorithm.

Input Noise and Amount of Noise added		SNR value	
		Proposed method using Modified cuckoo search	Existing method using Cuckoo search
Babble Noise	5dB	21.7874	10.5691
	10dB	33.2423	17.7956
	15dB	36.1727	24.7640
Car Noise	5dB	21.10028	10.7005
	10dB	31.2160	17.9095
	15dB	33.3417	24.7302
Exhibition Noise	5dB	20.4050	10.5710
	10dB	31.1152	17.8248
	15dB	33.2180	24.9196
Street Noise	5dB	20.0636	10.7593
	10dB	29.8668	17.7414
	15dB	31.8443	24.9270
Restaurant Noise	5dB	21.8038	10.6222
	10dB	34.2001	17.6489
	15dB	36.9347	24.5211

5. CONCLUSION

In this work, Modified cuckoo search based generation of weight is computed for noise suppression and improvement of speech signal is proposed. The input database noises include train noise, babble, car, exhibition hall, restaurant, street, and airport and train-station noise and it is processed using three stages. They are Feature extraction stage, Modified cuckoo search based weight generation in order to reduce the noises present in the input signal and ranking and

merging stage. Here, the Feature extraction is supported out using Amplitude magnitude spectrogram in which the classification of signals is done to generate the initial population. Then by using modified levy flight and Velocity in PSO the merging and ranking the best solution takes place. Thus the results obtained proved the effectiveness of the proposed technique and was able to suppress noise and improve the speech signal. Overall, the summary of finding using proposed method recommends that speech intelligibility can be improved by estimating the signal-to-noise ratio in each time-frequency unit.

REFERENCES

[1] Jaime Lorenzo-Trueba , Roberto Barra-Chicote , Rubén San-Segundo ,Javier Ferreiros , Junichi Yamagishi and Juan M. Montero, "Emotion transplantation through adaptation in HMM-based speech synthesis", Computer Speech and Language, No.34, pp. 292–307, 2015

[2] Seung Yun, Young-Jik Lee, and Sang-Hun Kim, "Multilingual Speech-to-Speech Translation System for Mobile Consumer Devices", IEEE Transactions on Consumer Electronics, Vol. 60, No. 3, August 2014

[3] Tomoki Toda, Mikihiro Nakagiri and Kiyohiro Shikano, "Statistical Voice Conversion Techniques for Body-Conducted Unvoiced Speech Enhancement", IEEE Transactions on audio, speech, and language processing, vol. 20, No. 9, November 2012.

[4] Vikramjit Mitra, Wen Wang, Andreas Stolcke, Hosung Nam, Colleen Richey, Jiahong Yuan, Mark Liberman, "Articulatory trajectories for large-vocabulary speech recognition",2013

[5] Hao-teng Fan ,Pao-han Lin and Jeih-weih Hung, "Noise Suppression based on nonnegative matrix factorization for robust speech recognition", 2014.

[6] Jindong Liu and Guang-Zhong Yang, "Robust speech recognition in reverberant environments by using an optimal synthetic room impulse response model", Speech Communication, No.67, pp. 65–77, 2015.

[7] Heikki Kallasjoki, Jort F. Gemmeke, and Kalle J. Palomäki, "Estimating Uncertainty to Improve Exemplar-Based Feature Enhancement for Noise Robust Speech Recognition", IEEE/ACM Transactions on audio, speech, and language processing, vol. 22, no. 2, February 2014.

[8] Vikramjit Mitra, Horacio Franco, Martin Graciarena and Arindam Mandal, "Normalized amplitude modulation features for large vocabulary noise-robust speech recognition",2012.

[9] Philip Harding and Ben Milner, "Reconstruction-based speech enhancement from robustacoustic features", Speech Communication, No.75, pp.62–75, 2015.

[10] Wenhao Yuan and Bin Xia, "A speech enhancement approach based on noise classification", Applied Acoustics, No.96, pp. 11–19, 2015

[11] Nasser Mohammadiha,, Paris Smaragdis, and Arne Leijon, "Supervised and Unsupervised Speech Enhancement Using Nonnegative Matrix Factorization", IEEE Transactions on audio, speech, and language processing, vol. 21, No. 10, 2013

[12] Anil Garg and O. P. Sahu, "Cuckoo Search BasedOptimal Mask Generation for Noise Suppression and Enhancement of Speech Signal", Journal of King Saud University - Computer and Information Sciences.

[13] Christian D. Sigg,, Tomas Dikk, and Joachim M. Buhmann, "Speech Enhancement Using Generative Dictionary Learning", IEEE Transactions on audio, speech, and language processing, vol. 20, No. 6, August 2012.

BIOGRAPHIES

Research Scholar, Dept. of ECE, T-John Institute of Technology, Bangalore-Karnataka , India

Professor and R&D Head, Dept. of ECE, Rajarajeswari Engineering College, Bangalore, Karnataka, India