International Journal of application or Innovation in engineering and management

ijaiem.com/Jan 2024/ Volume 13/Issue 1/Article No-1/44-48

ISSN: 2319-4847

Cuckoo search based optimal mask generation for noise suppression and enhancement of speech signal

Nagam Aanjaneyulu[1], Arekatla Madhava Reddy[2], Chevula Rekha[3],

Dr. G. Samba Siva Rao[4]

Associate Professor[1], Assistant Professor[2,3], Professor[4]

anji.amrexamcell@gmail.com[1],amreddy2008@gmail.com[2], rekhavenkat16@gmail.com[3],profgssrao@gmail.com[4]

Department of CSE, A.M. Reddy Memorial College of Engineering and Technology, Petlurivaripalem, Narasaraopet, Andhra Pradesh

Article Info

Received: 27-12-2023 Revised: 05 -01-2024 Accepted: 15-01-2024 Published:25/01/2024

Abstract

In this work, we offer an improved mask-based noise reduction strategy for improving voice signals. First, the Amplitude Magnitude Spectrogram is determined by decomposing the noisy speech signal into its component time-frequency (TF) components. (AMS). The signals are then sorted into groups according to their quality ratio to provide a preliminary set of answers. The Cuckoo search technique is then used to produce the best mask for each class. The waveform synthesis step next involves windowing the filtered waveforms, multiplying them by the best mask value, and summing the products to get the final, improved target signal. Experiments were run on many datasets to gauge the suggested method's efficacy, and the results were compared to those of established methods using signal-to-noise ratio. The acquired findings validated the efficiency of the suggested method in reducing background noise and improving the quality of the spoken signal.

KEYWORDS:

Noise cancellation; Speech signal amplification; Extracting AMS features; Cuckoo searching; Creating a waveform; Creating an optimal mask.

Introduction

Over the last several decades, researchers have paid a lot of attention to the issue of speech improvement. (Hu and Loizou, 2007). In particular, it is concerned with enhancing the effectiveness of voice communication systems in busy, chaotic settings like traffic and crowds. (Hong et al., 2009). There is a plethora of documented speech improvement techniques, from spectrum subtraction to subspace to statistical-model based and wiener type. (Hu and Loizou, 2007; Kim and Loizou, 2011). By deducting the average of the noise spectrum from the speech spectrum, spectral subtraction provides a rough approximation of the clean speech signal. (Boll, 1979). In the beginning, when there is no voice signal, the noise spectrum is estimated. (Boll, 1979). Intelligibility and signal-to-noise ratio (SNR) are common metrics used to evaluate the efficacy of speech augmentation techniques. (Kim and Loizou, 2011; Christiansen et al., 2010; Ma et al., 2010). Algorithms for assessing and enhancing intelligibility and SNR have been developed by a number of academics and professionals. (Hu and Loizou, 2007; Christiansen et al., 2010). Classic methods such as spectral subtraction, Wiener filtering, and maximum likelihood may be expressed as a function of this apriorism SNR (Loizou, 2006) to make decisions in voice enhancement and noise reduction. (Scalar and Filho, 1996). Although the apriorism SNR calculation might be helpful in real-time applications, the local SNR is usually more accurate and should be used instead. (Wolfe and Godsell, 2003). Using a weighted average of the

International Journal of application or Innovation in engineering and management

Ijaiem.com/Jan 2024/ Volume 13/Issue 1/Article No-1/44-48

ISSN: 2319-4847

prior SNR estimate and the current SNR estimate, Ephraim and Melah implemented the decision driven technique for SNR estimation. (Ephraim and Melah, 1984; Chen and Loizou, 2011). The modified decision-directed technique to calculating the gain function relies heavily on the posteriori and a priori SNRs. (Ephraim and Melah, 1984). Maximum a posterior (MAP) estimators Lu and Loizou have a gain function that is equal to the gain function utilized in the ideal binary mask for computational auditory scene analysis. (2011). Kim et al. also provided very important study.

Crazy-making search algorithms Look for cuckoos (CS) One of the most recent optimization techniques is based on the idea that obligatory brood parasitism of certain cuckoo species deposit their eggs in the nests of other host birds who are of different species (Yang, 2009; Valiant et al., 2011). Each cuckoo lays just one egg at a time and deposits it in a nest selected at random; this is only one of three idealized principles explored in Cuckoo Search. The second rule stipulates that the best nests, those producing the highest quality eggs, will be passed down to subsequent generations. The third rule specifies that the number of host nests is fixed, and the egg left by a cuckoo has a chance of being detected by the host bird between 0 and 1. The host bird might either discard the egg or leave the nest altogether to start a new one. Nests are presumed to be replaced by new nests at a fixed rate. A solution's quality or fitness for a maximizing issue may be directly correlated to the magnitude of the objective function. The algorithm is based on the Levy flying behaviour of some birds and fruit flies, as well as the obligatory brood parasite behaviour of certain species of cuckoos. The method uses Levy flight for updating, compares results using fitness functions, and makes appropriate replacements.

Generation of the best possible masks using cuckoo searches

An optimum mask generation module, a waveform synthesis module, and a feature extraction module (Kim et al., 2009) make up the strategy used in this research for noise suppression and voice improvement. In order to extract features from the speech signal, both the original and the noisy speech signal are input, and then the best mask is constructed using cuckoo search. The improved signal is then calculated by windowing the filtered waveforms, multiplying them by the appropriate mask value, and summing the results in the waveform synthesis module. Fig. 1 shows a block schematic of the suggested method.

Module for extracting features for this lesson

The Amplitude Magnitude Spectrogram (AMS) is used to extract characteristics from the input voice corpus (Kim et al., 2009). A combination of the clean speech signal and the noisy signal will make up the input speech signal. To prepare the input signal for future processing, it is first sampled, quantized, and then pre-emphasized. In Fig. 2, we see a block diagram of the process through which AMS features are extracted. Band pass filters are then used to do a TF (Time-Frequency) decomposition on the processed data. Band-pass filters are designed to only let through signals that fall within a certain frequency range, and this module (Kim et al., 2009) does just that by dividing the signals into 25 TF units, each of which contributes to a channel denoted by Cowher 1 6 I 6 25.

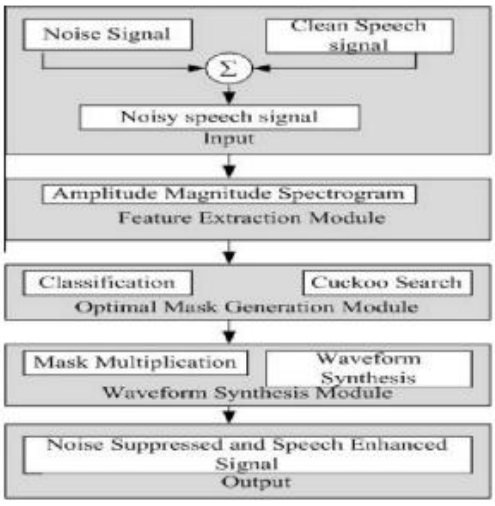

Figure 1 Block diagram of the proposed technique.

frequencies while suppressing undesired ones. Therefore, there will be transmissions within the range of frequencies specified for each of the 25 band channels under consideration. Here, the upper and lower frequency boundaries of each channel, Ui and Li, are established: Once the channel bands have been formed, the envelope of each band may be determined using full wave rectification. This envelope is then decimated by a factor of three before being divided into 128 samples of 32 ms with an overlap of 64 samples. (Lu and Loizou, 2011). Let's say the i-th channel forms a total of Ni segments, and we'll refer to each of those segments as Sij, where i=6, j=25, and Ni=the total number of segments. Henning windowing (Saliva Hanan, 2010) is used to the segmentation's sampled signals to filter out noise and bring out the signals' true peaks. Each channel's modulation spectrum, with a frequency resolution of 15.6 Hz, is obtained by first zero-padding the windowed signals and then performing a 256-point Fast Fourier Transform (FFT). (Kim et al., 2009).

Therefore, FFT is used to acquire the modulation spectrum for all 25 channels, and then the spectra are multiplied by 15 triangular-shaped windows that are evenly spaced over the 15.6-400 Hz range for each channel. (Kim et al., 2009). All of them are added together to form 15 amplitudes in the modulation spectrum, which are the AMS feature vectors. (Kim et al., 2009). When compared to more traditional feature extraction methods, AMS produces superior outcomes when applied to loud voice signals. This is the result of a concerted effort using the processes of segmentation, windowing, FFT, and multiplication by a triangle function. If we denote the time slot by / and the sub-band by k, then we may write the feature vector as AFk;/. (Kim et al., 2009). To account for the possibility of subtle shifts in the time and frequency domains, we additionally include delta functions into the retrieved features. Kim et al. (2009) provides the following time delta functions DAT:

$$\Delta A_T(\lambda, \phi) = A_F(\lambda, \phi) - A_F(\lambda, \phi - 1), \text{ where } \phi = 2, \ldots, T$$

The frequency delta function DAS is as given below:

$$\Delta A_s(\lambda, \phi) = A_F(\lambda, \phi) - A_F(\lambda - 1, \phi) \text{ where } \lambda = 2, \ldots, B$$

The overall feature vector Aðk;/Þ including the delta functions can be defined as

$$A(\lambda, \phi) = [A_F(\lambda, \phi), \Delta A_T(\lambda, \phi), \Delta A_S(\lambda, \phi)]$$

Therefore, we have used AMS feature extraction on a massive speech signal corpus to get the features. (Kim et al., 2009).

Discussions on Experimental Findings

The suggested method for voice improvement and noise reduction runs on a machine with 4 GB RAM, 32 bit software, and an i5 processor, and is implemented in MATLAB Version 2012 with COLEA (Kim et al., 2009). In Section 4.1, we describe the dataset, and in Section 4.2, we provide the experimental findings.

Databank Synopsis

The experimental database was collected from the Loizou database provided in Kim et al. (2009). The database was built to facilitate the evaluation of speech enhancement strategies. Thirty IEEE words were included in the noisy database after being corrupted by eight distinct types of real-world disturbances of varying signal-to-noise ratios. The sounds were gathered from the AURORA database (Hirsch and Pearce, 2000) and include things like traffic, conversations, aviation traffic, and more. Tucker Davis Technologies (TDT) recording equipment was used to capture the IEEE sentence database in a soundproof room. Three men and three women each read the sentences. The speech was captured at 25 kHz and then down sampled to 8 kHz.

Consequences of Experiments

Figure 6 displays the simulation's output, which consists of plots of the input signal, the noisy signal, and the de-noised signal. Between 0 and 2.5 kHz, the signal power is shown against its associated frequency. 0 decibels, 5 decibels, 10 decibels, and 15 decibels of chatter noise, automobile noise, exhibition noise, restaurant noise, street noise, and train noise were utilized as maskers. Participants were exposed to 24 different permutations (0 SNR, 5 SNR, 10 SNR, 15 SNR) over 4 SNR levels and 6 masker types. The acquired findings validated the efficiency of

the suggested method in reducing background noise and improving the quality of the spoken signal. Figure 8 is a graphical illustration of the % improvement in SNR at a 10 dB level for different maskers.

Conclusions drawn from a comparison of (tables 1 and Figs. 7 and 8)

Using the Bayesian Classifier as a benchmark, we compared the suggested method with the usual assessment measure of SNR. Various

Table 2 SSNR for different cases.

Noise level (dB)	Babble noise		Car noise		Exhibition noise		Restaurant noise		street noise		Train noise	
	Proposed SSNR	Bayesian SSNR	Proposed SSNR	Bayesian SSNR	Proposed SSNR	Bayesian SSNR	Proposed SSNR	Bayesian SSNR	Proposed SSNR	Bayesian SSNR	Proposed SSNR	Bayesian SSNR
0	−4.55	−7.13	−5.05	−7.68	−4.88	−7.46	−4.54	−7.06	−4.75	−7.64	−4.52	−7.35
5	−1.80	−5.39	−2.33	−5.40	−1.19	−4.99	−2.23	−5.20	−1.07	−5.28	−1.84	−5.43
10	1.09	−4.82	0.77	−4.83	0.96	−4.87	0.93	−4.75	1.55	−4.62	0.69	−4.78
15	4.17	−3.00	3.45	−3.16	3.70	−3.13	4.33	−2.95	4.61	−3.10	4.31	−3.07

Noise sources included chatter, trains, automobiles, displays, eateries, and the streets. Noise at 0 dB, 5 dB, 10 dB, and 15 dB levels have been studied in all scenarios. The average SNR for both the suggested method and the Bayesian approach is shown in Fig. 7. The efficacy of the suggested method is shown by the fact that it achieves better outcomes than Bayesian. When comparing the suggested method to the Bayesian method, the best SNR value achieved is 31.0977 dB. When compared to the Bayesian method's SNR value of 10.78 dB, the suggested methods averaged out to 16.79 dB. Figure 8 displays the % SNR improvement for a 10 dB reduction in noise. Better performance for the suggested method was achieved by using an optimum mask. The Segmental signal-to-noise ratio (SSNR) calculation also requires the mask value; therefore, its significance cannot be overstated. Here, the method utilizes subsegments of both the target and masker signals. The system then calculates segment energies, followed by SNRs, and finally provides the average segmental SNR. (dB). You may compare the Segmented SNR between the suggested method and the Bayesian one in Table 2. The results show that the suggested method has improved upon the SSNR values. When compared to the Bayesian method's net average SSNR of -5.31, the suggested method's SSNR was around 0.02.

Conclusion

In this work, we provide a cuckoo search-based method for generating optimum masks to reduce background noise and improve the quality of a voice signal. The method may be broken down into its three constituent parts: feature extraction, optimum mask creation, and waveform synthesis. In order to populate the cuckoo search algorithm with initial candidates, feature extraction is performed using AMS, and signals are classified. The suggested method was simulated using a number of data sets. The SNR metric allowed for comparison with earlier methods as well. The acquired findings validated the efficiency of the suggested method in reducing background noise and improving the quality of the spoken signal. The best SNR achieved with the suggested method is 31.0977 dB, whereas the best SNR achieved with the Bayesian method is 24.67 dB. When compared to the Bayesian method's SNR value of 10.78 dB, the suggested methods averaged out to 16.79 dB. The suggested method yielded significant improvements in intelligibility with a minimum of training data. Results reveal that measuring the signal-to-noise ratio in each time-frequency unit may increase speech intelligibility, as a whole, when utilizing the suggested method.

References

[1] Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process 27, 113– 120.

[2] Brungart, D., Chang, P., Simpson, B., Wang, D., 2006. Isolating the energetic component of speech-on-speech masking with ideal timefrequency segregation.

[3] J. Acoust. Soc. Amer. 120, 4007–4018. Chen, F., Loizou, C., 2011. Impact of SNR and gain function over – and under-estimation on speech intelligibility. Speech Commun. 54, 272–281.

[4] Chirstiansen, C., Pedersen, M.S., Dau, T., 2010. Prediction of speech intelligibility based on an auditory preprocessing model. Speech Commun. 52, 678–692.

[5] Ephraim, Y., Malah, D., 1984. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. Acoust Speech Signal Process ASSP-32 (6), 1109–1121.

[6] Hirsch, H., Pearce, D., 2000. The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions. ISCA ITRW ASR, September 18–20.

[7] Hong Y.L., Qing H.Z., Guang L.R., Bao J.X., Speech Enhancement algorithm Based on Independent Component Analysis. 5th IEEE International Conference on Natural Computation, 2009, pp. 598– 602.

[8] Hu, Y., Loizou, P., 2007. Subjective comparison of speech enhancement algorithms. Speech Commun. 49, 588–601.

[9] Kim, Gibak, Loizou, Philipos C., 2010. Improving speech intelligibility in noise using environment optimized algorithms. IEEE Trans. Audio Speech Lang. Process. 18 (8), 2080–2090.

[10] Kim, G., Loizou, C., 2010. A new binary mask based on noise constraints for improved speech intelligibility. Interspeech, Chiba, Japan, 1632–1635.

[11] Kim, Gibak., Loizou, Philipos C., 2011. Reasons why speechenhancement algorithms do not improve speech intelligibility and suggested solutions. IEEE Trans. Audio Speech Lang. Process. 19 (1), 47–56.

[12] Kim, Gibak, Yang, Lu, Yi, Hu, Loizoua, Philipos C., 2009. An algorithm that improves speech intelligibility in noise for normalhearing listeners. J. Acoust. Soc. Am. 126 (3), 1486–1492.

[13] Li, N., Loizou, P.C., 2008. Factors influencing intelligibility of ideal binary-masked speech: implications for noise reduction. J. Acoust. Soc. Amer. 123 (3), 1673–1682.

[14] P.C. Loizou, 2006, Speech processing invocoder-centric cochlear implants, In: Møller, A.R. (Ed.), Cochlear and Brainstem Implants, Advances in Oto- Rhino-Laryngology, Karger, Basel, Switzerland, 64, pp. 109–143.

[15] Loizou, P.C., 2007. Speech Enhancement: Theory and Practice. CRC Press. Youyi, Lu, Cooke, Martin, 2009. The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. Speech Commun. 51, 1253–1262.

[16] Lu, Y., Loizou, P., 2011. Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty. IEEE Trans. Audio Speech Lang. Process. 19 (5), 1123–1137.

[17] Jianfen, Ma, Loizou, P.C., 2010. SNR loss: a new objective measure for predicting the intelligibility of noise-suppressed speech. Speech Commun. 53, 340–354.

[18] Mandal, Sangeeta, Ghoshal, Sakti Prasad, Kar, Rajib, Mandal, Durbadal, 2012. Design of optimal linear phase FIR high pass filter using craziness based particle swarm optimization technique. J. King Saud Univ. Comp. Inform. Sci. 24 (1), 83–92.

[19] Muhammad, Ghulam, 2010. Noise-robust pitch detection using autocorrelation function with enhancements. J. King Saud Univ. – Comp. Inform. Sci. 22, 13–28.

[20] Salivahanan, Gnanapriya, 2010. Digital signal processing, second ed. Tata McGraw Hill. Scalart, P., Filho, J.V., 1996. Speech enhancement based on apriori signal to noise estimation, In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2. IEEE, pp. 629–632.