Journal of King Saud University – Computer and Information Sciences (2015) 27, 269–277

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

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

Anil Garg *, **O.P. Sahu**

Department of ECE, NIT Kurukshetra, India

Received 25 April 2013; revised 7 March 2014; accepted 3 April 2014
Available online 18 June 2015

KEYWORDS

Noise suppression;
Enhancement of speech signal;
AMS feature extraction;
Cuckoo search;
Waveform synthesis;
Optimal mask

Abstract In this paper, an effective noise suppression technique for enhancement of speech signals using optimized mask is proposed. Initially, the noisy speech signal is broken down into various time–frequency (TF) units and the features are extracted by finding out the Amplitude Magnitude Spectrogram (AMS). The signals are then classified based on quality ratio into different classes to generate the initial set of solutions. Subsequently, the optimal mask for each class is generated based on Cuckoo search algorithm. Subsequently, in the waveform synthesis stage, filtered waveforms are windowed and then multiplied by the optimal mask value and summed up to get the enhanced target signal. The experimentation of the proposed technique was carried out using various datasets and the performance is compared with the previous techniques using SNR. The results obtained proved the effectiveness of the proposed technique and its ability to suppress noise and enhance the speech signal.

1. Introduction

The problem of speech enhancement has received a significant amount of research attention over the past several decades (Hu and Loizou, 2007). Particularly, it focuses on improving the performance of speech communication system in noisy environments such as traffic and crowd (Hong et al., 2009). Many speech enhancement algorithms such as spectral subtraction, subspace, statistical-model based and wiener type have been reported (Hu and Loizou, 2007; Kim and Loizou, 2011). Spectral subtraction is based on principle of obtaining the estimate of clean speech signal by subtracting the average of noise spectrum from noisy speech spectrum (Boll, 1979). The noise spectrum is estimated initially in the absence of speech signal (Boll, 1979). The performance of the speech enhancement algorithms is usually measured in terms of intelligibility and signal-to-noise ratio (SNR) (Kim and Loizou, 2011; Chirstiansen et al., 2010; Ma et al., 2010). Several researchers and professionals have developed various algorithms for estimating and improving intelligibility and SNR (Hu and Loizou, 2007; Chirstiansen et al., 2010). In many speech enhancement and noise reduction algorithms, the decision is based on the apriori SNR (Loizou, 2006), and the classic algorithms like spectral subtraction, Wiener filtering, and maximum likelihood, can be formulated as a function of this

* Corresponding author.
E-mail addresses: anilgarg0778@gmail.com, agarg001@yahoo.com (A. Garg).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2014.04.006
1319-1578

a priori SNR (Scalart and Filho, 1996). In real-time applications, the apriori SNR estimation is useful, but in the ideal situation the local SNR is preferable instead of the apriori SNR (Wolfe and Godsill, 2003). For example, Ephraim and Malah used the decision directed approach for signal-to noise ratio estimation by using the weighted average of the past SNR estimate and the present SNR estimate (Ephraim and Malah, 1984; Chen and Loizou, 2011). The posteriori and a priori SNRs are main function for computing gain function using modified decision-directed approach (Ephraim and Malah, 1984). The gain function used in ideal binary mask for computational auditory scene analysis is identical to the gain function of the Maximum a posterior (MAP) estimators Lu and Loizou (2011). Another significant research was presented by Kim et al. (2009) and Kim and Loizou (2010), where the input signals were broken down into time–frequency units and the features were extracted by the AMS feature extraction technique. In this approach, binary decisions (weight value zero or one) were taken based on the Bayesian classifier, as to whether each T–F unit is dominated by the target or the masker. These speech enhancement algorithms/approaches have been reported to estimate the original speech, degraded by various types of noises (Lu and Loizou, 2011; Kim et al., 2009; Kim and Loizou, 2010; Muhammad, 2010). However, the degree of improvement, measured in terms of intelligibility and SNR, is not easy (Kim and Loizou, 2011; Chirstiansen et al., 2010; Ma et al., 2010). This is primarily due to lack of good estimation of the noise spectrum, especially when it is non stationary (Kim and Loizou, 2011). However, a high signal–to-noise ratio is always desirable to increase speech intelligibility (Kim and Loizou, 2011; Chirstiansen et al., 2010; Ma et al., 2010). In recent studies, the binary mask (Kim and Loizou, 2010) retains the time–frequency (T–F) regions where the target speech dominates the masker (noise) (e.g., local SNR > 0 dB) and removes T–F units where the masker dominates (e.g., local SNR < 0 dB) (Kim and Loizou, 2010). Although, speech produced in the presence of noise called ‘‘Lombard speech’’ has been found to be easily understandable than speech produced during silence (Lu and Cooke, 2009). In previous studies, large gain in intelligibility can be obtained by multiplying the noisy signal with the ideal binary mask signal, even at extremely low (5, 10 dB) SNR levels (Brungart et al., 2006; Li and Loizou, 2008). Kim et al. (2009) and Kim and Loizou (2010) presented the generation of binary mask with the help of Bayesian classifier technique that is lazy classification technique. Since the classification with the lazy classifier, the generation of binary mask will not be an optimal one. If the binary mask is not an optimal one, it will affect the performance of the speech enhancement. This paper presents optimal mask generation using cuckoo search algorithm (Yang, 2009) which is a kind of optimization algorithm (Mandal, 2012; Venkata Rao and Waghmare, 2014) for speech enhancement to improve the SNR and thus intelligibility. The proposed algorithm optimizes the masking parameters in order to suppress the noise effectively for enhancement of speech signal. Comparison and simulation results of our proposed method are better in terms of SNR than the Bayesian classifier technique.

The rest of the paper is organized as follows: A brief description of Cuckoo search algorithm is given in Section 2. The cuckoo search based optimal mask generation is explained in Section 3. The simulation results and discussions are presented in Section 4. The paper is concluded in Section 5.

2. Cuckoo search algorithms

Cuckoo search (CS) Yang, 2009; Valian et al., 2011 is one of the latest optimization algorithms and was developed from the inspiration that the obligate brood parasitism of some cuckoo species lay their eggs in the nests of other host birds which are of other species. In Cuckoo Search, three idealized rules are considered which say that each cuckoo lays one egg at a time, and dumps its egg in a randomly chosen nest. The second rule states that best nests with high quality of eggs will carry over to the next generations and the third one says that the number of available host nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with a probability in the range 0–1. In this case, the host bird can either throw the egg away or abandon the nest, and build a completely new nest. It is also assumed that a definite fraction of the nests are replaced by new nests. For a maximization problem, the quality or fitness of a solution can simply be proportional to the value of the objective function. The algorithm is based on the obligate brood parasitic behavior of some cuckoo species in combination with the Levy flight behavior of some birds and fruit flies.

In the algorithm, updation is carried out using Levy flight and comparison is made with the use of fitness functions and suitable substitutions are made. Levi flight is carried out on ym_i to yield to get a new cuckoo ym_i^* which is given by: $ym_{i1}^* = ym_{i1}^{(t+1)} = ym_{i1}^{(t)} + \Delta \otimes Levy(y)$, where the levy sharing is specified by: $Levy(y) = \sqrt{\frac{c}{2\pi}} \cdot \frac{e^{-\frac{1}{2}\left(\frac{c}{y}\right)}}{y^{3/2}}$, where c is arbitrary constant. Consequently, some other nest is observed and its fitness function is found out. If the fitness of the Levy flight made nest is superior to the fitness of the nest in consideration, then substitute nest signal values by the host nest Levy performed values. For each iteration, a portion of the utmost horrifying nests are done away with and fresh nests are constructed as replacement.

Based on the above mentioned rules, the basic steps of the Cuckoo search can be summarized as the pseudo code as follows (Yang, 2009; Valian et al., 2011):

Pseudo code:
Objective Function: Maximize the SNR ratio and to obtain the optimal mask weight for each class
Start
For every class Cl_i for $0 < I \leqslant 3$ perform:
The initial population of the class cl_i in consideration is $G_i = \{g_{i1}, g_{i2} \ldots\ldots\ldots g_{iN_{ci}}\}$
Generate 25 host nests $H = \{h_1, h_2 \ldots\ldots\ldots\ldots h_{25}\}$ and consider the signals $Y_i = \{y_{i1}, y_{i2} \ldots\ldots y_{iN_h}\}$ in the i^{th} host nest for $0 < i \leqslant 25$
While (stop criteria)
Perform the levy flight $y_{i1}^* = y_{i1}^{(t+1)} = y_{i1}^{(t)} + \Lambda \otimes Levy(x)$ for all signals in the i^{th} host nest
Find the fitness of the new solution F_i where fitness is the SNR ratio
Choose another random nest j and find the fitness value F_j
If ($F_i > F_j$)