Int. J. Electron. Commun. (AEÜ) 72 (2017) 125–133

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

journal homepage: www.elsevier.com/locate/aeue

An immunological approach based on the negative selection algorithm for real noise classification in speech signals

Caio Cesar Enside de Abreu [a,b], Marco Aparecido Queiroz Duarte [c], Francisco Villarreal [d]

[a] State University of Mato Grosso - UNEMAT, Department of Computing, Santa Rita Street 128, Alto Araguaia, Brazil
[b] São Paulo State University - UNESP, Department of Electrical Engineering, Brasil Avenue 56, Ilha Solteira, Brazil
[c] State University of Mato Grosso do Sul - UEMS, Department of Mathematics, Rodovia MS 306 KM 06, Cassilândia, Brazil
[d] São Paulo State University - UNESP, Department of Mathematics, Brasil Avenue 56, Ilha Solteira, Brazil

ARTICLE INFO

Article history:
Received 6 April 2016
Accepted 6 December 2016

Keywords:
Noise classification
Speech enhancement
Artificial immune systems
Negative selection algorithm
Dual-tree complex wavelet transform

ABSTRACT

This paper presents a new approach to detect and classify background noise in speech sentences based on the negative selection algorithm and dual-tree complex wavelet transform. The energy of the complex wavelet coefficients across five wavelet scales are used as input features. Afterward, the proposed algorithm identifies whether the speech sentence is, or is not, corrupted by noise. In the affirmative case, the system returns the type of the background noise amongst the real noise types considered. Comparisons with classical supervised learning methods are carried out. Simulation results show that the artificial immune system proposed overcomes classical classifiers in accuracy and capacity of generalization. Future applications of this tool will help in the development of new speech enhancement or automatic speech recognition systems based on noise classification.

1. Introduction

Speech processing systems are essential in many branches of telecommunications, as well as in the entertainment industry. Examples include applications on signal encoding, automatic speech recognition, mobile applications, communication in noisy environments and medicine. In the specific case of speech enhancement, the task consists of extracting the original signal from the corrupted speech signal [1]. Classical speech enhancement methods include spectral subtraction and power spectral subtraction, minimum mean square error (MMSE) estimators based, Wiener filtering and wavelet thresholding. Regardless of the fact that these methods have their own theoretical basis with specific objectives, all of them require noise estimation in order to perform noise reduction. In fact, the better the noise estimation the greater the denoising speech quality [2,3].

For mobile communication systems, background noise is the main drawback. Noise can damage speech communication quality and the performance of automatic speech recognition algorithms [4]. According to [5], the design of speech enhancement algorithms usually does not take into account the differences in statistical properties of different noise types. It can be the cause of failure of some algorithms in specific noise environments, as can be seen in [6]. In this sense, the development of ever more powerful devices, such as smartphones, tablets and hearing aids, enables the design of algorithms which work efficiently in any noisy environment, by incorporating noise classification. As an example of real-time implementation, in [7], an implementation for smartphones that includes frequency domain transformation, noise classification and suppression was presented.

One of the first speech enhancement techniques capable of practical applications is spectral subtraction (SS), proposed by Boll in [1]. The basic idea of SS is to estimate the noise frequency spectrum and then subtract it from noisy speech. The SS technique works well when the noise frequency spectrum is uniform or when noise is stationary. Under the conditions of real noise environments, SS generates undesirable tones in the processed signal, known as "Musical Noise". However, musical noise is not a specific shortcoming of SS. Wiener filtering [8], statistical-model approaches such as MMSE estimators [2,9], log-spectral amplitude estimators [10] and MAP estimators [11] are also affected by this problem. Scarlat and Vieira Filho, in [12], suggested that *a posteriori* signal-to-noise ratio (postSNR) and *a priori* signal-to-noise ratio (prioSNR) concepts could avoid or attenuate the musical noise problem by means of a derived Wiener filter. Simulations in [12,3] showed improved results; however, for some noise environments the performance was not successful. This was due to

E-mail addresses: caioenside@unemat.br (Caio Cesar Enside de Abreu), marco@uems.br (M.A.Q. Duarte), villa@mat.feis.unesp.br (F. Villarreal)

http://dx.doi.org/10.1016/j.aeue.2016.12.004
1434-8411/

prioSNR estimation by the decision-directed method, which depends fundamentally on good noise estimation [2].

Wavelet based methods are also susceptible to poor performance due to noise estimation. The wavelet thresholding approach accepts as noise coefficients those with absolute value under a certain value (threshold), after which such coefficients are modified. Originally, threshold estimation, proposed by Donoho in [13], was based on the assumption that background noise is white Gaussian. Therefore, results are not satisfactory in real noise conditions. Several strategies to improve original threshold estimation were proposed. Examples are adaptive thresholding [14] and new statistical models to threshold computation [15–17].

As can be seen, it is natural that future research in speech enhancement focuses on the treatment and understanding of background noise. In this sense, the development of systems that are able to predict the noise type present in a noisy sentence are indispensable to statistical methods or for obtaining optimal parameters in algorithms. This tendency is confirmed in two current works. In [5], authors have used noise classification to choose optimal smoothing parameters in noise and prioSNR estimation. Such optimal parameters were used in the log-spectral amplitude speech estimator algorithm. In [18], noise classification was employed to predict the background noise and a specific weighted denoising auto-encoder model, [19], was used in the estimation of the clean power spectrum, employed in the implementation of a Wiener filter. Besides these applications in speech enhancement, noise classification has been used in automatic speech recognition systems [20,21], acoustic scene classification [22,23] and hearing aid applications [24,25].

In [22], authors inserted noise classification for context-aware applications and a hidden Markov model based noise classifier was proposed. Without any distinction between speech presence or absent, mel-frequency cepstral coefficients were computed and used as input feature for the classifier. However, authors highlighted that noise-only segments would suitable for noise classification.

The noise classification method proposed in [5] is based on the traditional support vector machine (SVM) classifier. Features are acquired by mapping noise energy from the 256-point short time Fourier transform to the Bark domain. In other words, noise energy in each Bark band is calculated and used as an input feature for the classifier. In order to perform noise classification in a noisy speech signal, the first 15 frames are assumed to be noise-only segments. Classification is carried out only in these frames in the following way: within the 15 frames, noise is classified frame-by-frame and the noise type with greater vote number is selected to be the noise type in the whole sentence.

In a similar way, in [18], authors used the normalized subband power spectrum of noisy speech as an input feature from a classifier. However, in that case, a Gaussian mixture model is used as classifier [26]. As well as in [5], the first 10 frames of the noisy speech are assumed as noise-only segments and the noise classification is carried out frame-by-frame. The noise type with greater vote number is selected for classification. Furthermore, in order to address possible changes of noise type, the authors have used a voice activity detection (VAD) algorithm and noise classification performed every time that a speech absent frame is identified.

The motivation of this paper is the proposition of a methodology for real noise classification in speech signals in a frame-by-frame way, which will contribute to the scientific development of speech enhancement and other speech processing systems. Unlike the methods proposed in [5,18], for the proposed algorithm initialization a single noise-only frame is required. Furthermore, it is proposed in this paper that the system identifies whether the speech sentence is clean, or the noise level is so low that no processing (e.g. enhancement) is required. Thus, the proposed algorithm is easily coupled to other speech processing systems. Besides the classifiers used in the above noise classification methods, classifiers commonly used in pattern recognition problems include neural networks (NNs) [27] and decision trees (DTs) [28]. Therefore, simulations encompassing both classifiers are presented. Another motivation of this work is the inclusion of artificial immune systems (AIS) in the related area. As outlined previously, noise classification has become a tendency for speech processing, making possible the development of algorithms that will operate in a specific way for each real noise environment.

The main objective of this work is the development of an intelligent system for real noise classification based on an AIS [29,30]. AISs constitute a relatively recent approach in the artificial intelligence field. Researchers in the AIS field look to the biological immune system (BIS) for inspiration on how to solve problems in engineering and computer science [29].

Composed by a set of organs, cells and molecules, BIS aims to protect an individual from infections, eliminating foreign substances [31]. In a general way, it is able to recognize common structures from different classes of microorganisms, in order to generate an immune response. The exposure of the BIS to a foreign antigen increases its ability to respond more quickly to a new exhibition to the same antigen, characterizing the immunological memory concept [31,32].

From such biological concepts, several algorithms were proposed. Forrest et al., in [33], proposed an approach based on the generation of T-cells in the immune system which contributed with the broad dissemination of AISs [29]. In that paper, the authors proposed a method called negative selection algorithm (NSA) that was then applied to the problem of computer virus detection. Although originally applied to computer security, AISs have been applied to several areas; some examples are: pattern recognition [34–36], data mining [37–39], optimization [40–42], fault and anomalies detection [43,44], as well as machine learning [45].

This paper presents a real noise classifier for speech processing systems, based on NSA [33]. Among its characteristics, NSA is attractive for its simplicity of implementation and high accuracy in pattern recognition [29]. NSA is mainly based on simple comparisons between patterns by means of an affinity measure (e.g. a distance measure), unlike SVM and NNs with implementations based on optimization algorithms. In addition, unlike the works in [5,18], where frequency domain extracted features are used, the present work proposes a multiscale analysis from the dual-tree complex wavelet transform (dual-tree ℂWT) [46]. The dual-tree ℂWT is a relatively recent architecture developed to implement a complex wavelet transform first inserted in the speech enhancement context by Abreu et al. in [47].

From the speech signal, features are extracted by the dual-tree ℂWT and NSA is applied in order to detect whether a speech sentence is degraded, or not, by noise. In the affirmative case, extracted features are stored in a database. This procedure is repeated as often as necessary, until a set of detectors (antibodies) is built for each type of real noise environment considered, which is accomplished in off-line mode. In on-line mode, the proposed system will act by identifying and classifying background noise in speech signals. This procedure is inspired by the monitoring phase of the negative selection algorithm [33]. If the speech sentences do not show degredation, the monitoring phase will not be started.

The remainder of this paper is organized as follows: Section 2 presents NSA in its original form; proposed methodology is described in Section 3; Section 4 presents simulation details, including the features extraction process and results; and, finally, concluding remarks are presented in Section 5.