Applied Acoustics 132 (2018) 26–32

Contents lists available at ScienceDirect

Applied Acoustics

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

Real-time hierarchical classification of sound signals for hearing improvement devices

Fatemeh Saki, Nasser Kehtarnavaz*

Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA

A R T I C L E I N F O

Keywords:
Real-time hierarchical classification of sound signals
Computationally efficient classification of sound signals
Sound classification in hearing improvement devices

A B S T R A C T

This paper presents a real-time hierarchical approach to sound signal classification for utilization in hearing improvement devices. The developed classification hierarchy consists of three levels to classify speech, music and different noise types. A distinguishing attribute of this hierarchical approach is that effective features are computed as needed at different levels of the hierarchy making the classification process computationally efficient. This approach is compared to the conventional one-step classification approach by examining both trained and non-trained sound signals. The results obtained show higher classification rates as well as higher computational efficiency of this hierarchical approach compared to the conventional one-step approach.

1. Introduction

Classification of sound signals plays a major role in hearing improvement devices. Examples of such devices that benefit from a sound classification component or subsystem include hearing aids, cochlear implants, and smart headphones. It is well established that the hearing sensation of hearing aid users degrades considerably in noisy environments. Thus, attempts have been made at developing speech enhancement/noise reduction algorithms that are adaptive to different sound environments, e.g. [1–5]. Examples of commercially available hearing improvement devices that include a sound classification component are Phonak Bolero V hearing aid [6] and Cochlear Limited Necleus6 cochlear implant [7]. Classification of sound signals enables adapting the speech enhancement/noise reduction algorithm in such devices to different sound environments in an automatic manner.

In [8–12], our research team developed a number of approaches for the classification of noise signals. This work involves an extension of our previous classification approaches and is aimed at providing a real-time supervised classification component for hearing improvement devices in order to cope with three major categories of environmental sounds that are commonly encountered on a daily basis. These categories include speech in the presence or absence of background noise, music, and different types of background noise.

It is worth mentioning that recently deep learning approaches have been considered for sound classification purposes, for example [13–15]. However, their computational complexity remains a challenge for real-time deployment on resource limited hearing improvement devices that have limited processing power and memory. Our thrust in this paper is on the real-time aspect of a sound classification approach that can run in real-time on resource limited hearing improvement devices.

A typical sound signal classification component or subsystem possesses two major modules: feature extraction and classifier. Different types of environmental sound signals have been considered in the literature for different applications. As the number of features is increased, the computational complexity of the classification is increased. It is well-known that the combination of many individual features does not necessarily lead to higher classification rates and often causes a limitation as far as the real-time implementation aspect is concerned. Thus, from a practical implementation standpoint, it would be helpful to break down the above multi-class classification problem into several two-class classifiers, similar to the modulation signal classification discussed in [16]. This way, the classification process can be made computationally efficient by using a small number of effective features for each two-class classifier. This is achieved by performing the classification in a hierarchical manner. In this paper, a supervised hierarchical classification approach is developed to gain computational efficiency for real-time utilization in hearing improvement devices noting that a hierarchical approach avoids unnecessary computation of features and classifiers.

The rest of the paper is organized as follows. Section 2 discusses the developed real-time hierarchical classification approach. Then, in section 3, the details of the features used at each level of the hierarchy are mentioned. The experimental results and the real-time aspects appear in section 4. Finally, the conclusion is stated in section 5.

* Corresponding author.
E-mail address: kehtar@utdallas.edu (N. Kehtarnavaz).

https://doi.org/10.1016/j.apacoust.2017.11.007
Received 24 August 2017; Received in revised form 9 November 2017; Accepted 10 November 2017
Available online 21 November 2017
0003-682X/

2. Hierarchical sound signal classification

The hierarchical approach to classification of sound signals has been limited to a few studies in the literature. In [17], a hierarchical rule-based approach was developed to classify audio signals from movies or TV programs. In [18], a hierarchical approach was proposed for recognition of environmental noise events, where input sound signals were initially classified into road vehicle/non-road vehicle noise. This was then followed by additional classifiers to separate the noise classes of road vehicle in car, truck and motorbike, and the noise classes of aircraft, train and industrial machinery. The features considered were MPEG-7, mel-frequency cepstral coefficients (MFCC) and the classifiers used consisted of a k-nearest neighbor, a neural network, and a Gaussian mixture model (GMM) classifier. In [19], audio signal classification for a movie video abstraction scheme was presented in these three stages: (i) silence or environmental noise detection, (ii) speech and non-speech classification, and (iii) pure music/songs and speech with background music classification. In [20], a hierarchical algorithm for classifying urban mechanical sound signals consisting of aircraft, motorcycle, car, crowd, thunder, wind, train, and horn was covered where at the top level, sound signals were classified into two categories of mechanical and non-mechanical sounds, then at a lower level, the mechanical sound signals were classified into aircraft, motorcycle, car and train, and the non-mechanical sounds were classified into crowd, thunder, wind, and horn via GMM and Hidden Markov model (HMM) classifiers.

The focus of this work is on the development of a real-time or computationally efficient hierarchical classification approach for classifying non-quiet sound environments into music, speech, and background noise, where background noise signals are also classified into three major types of noise: stationary (e.g., machinery), semi-stationary (e.g., driving car) and non-stationary (e.g., babble). Hence, basically a five-class classification problem is addressed here in a hierarchical manner for the purpose of utilizing it in hearing improvement devices including hearing aids and cochlear implants.

Fig. 1 shows a block diagram of the developed hierarchical classifier. An incoming sound signal is first checked to see whether the environmental condition is quiet or not. No further processing is done for the quiet condition. Non-quiet sound signals are passed through a classifier to separate speech activity from absence of speech. If speech activity is not detected, the signal is passed to the second level of the hierarchy to see whether it is music or background noise. If the signal is detected to be noise, at the third level of the hierarchy, it is classified into three different noise types having stationary, semi-stationary and non-stationary statistical characteristics.

It is important to note that despite the existing hierarchical approaches that compute all the features at the start of the classification, in this hierarchical approach, effective features are extracted as needed at the appropriate level of the hierarchy. This approach allows the real-time operation of the classifier. It should be noted that the features that are already computed at higher levels are also utilized at lower levels. In the next section, the features that have been found effective at different levels of the hierarchy are described.

3. Features at different levels of hierarchy

A list of the effective features used at different levels of the hierarchy is provided in Table 1.

3.1. Quiet/ non-quiet condition

As shown in Fig. 1, an incoming sound signal is first seen to be quiet or non-quiet. This is achieved based on Sound Pressure Level (*SPL*). The louder a sound signal becomes, the greater the change in air pressure gets. SPL is computed as follows:

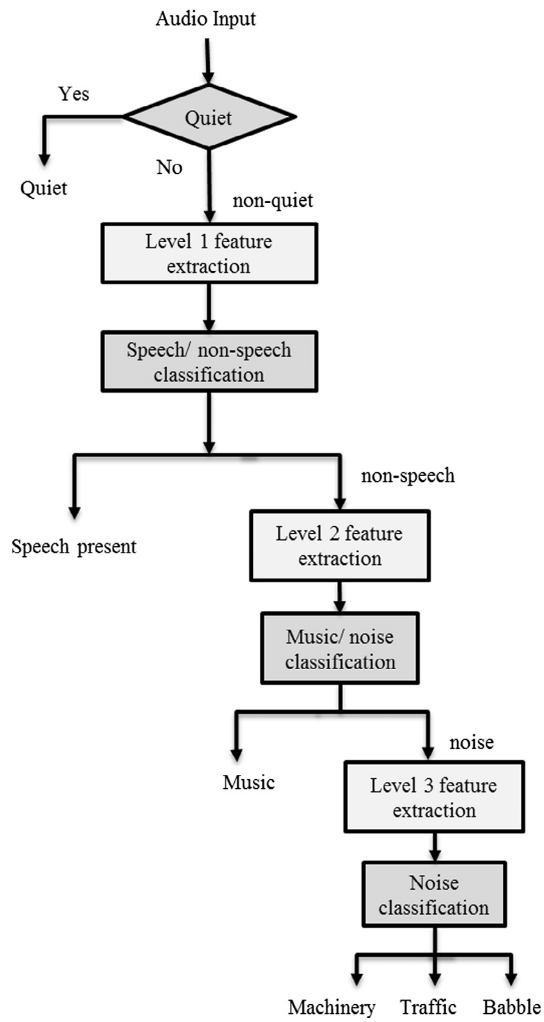

Fig. 1. Hierarchical classification of sound signals for hearing improvement devices.

Table 1
List of features at different levels of the hierarchical classification approach.

Classification levels	Features
Level 0: quiet/non-quiet	Sound Pressure Level (*SPL*)
Level 1: speech present/speech absent	Band-Periodicity $BP_b, b=1,...,B$, Subband Power Spectral Deviation $SPSD_b, b=1$, Spectral Centroid SC, High Zero-Crossing Rate Ratio (*HZCRR*), Low Short-Time Energy Ratio (*LSTER*), and Spectrum Flux
Level 2: music/noise	Spectral Centroid SC, High Zero-Crossing Rate Ratio (*HZCRR*), Low Short-Time Energy Ratio (*LSTER*), Spectrum Flux, Band-Periodicity $BP_b, b=1$, Band-Periodicity Deviation $BPD_b, b=1,...,B-1$, Subband Short-Time Energy Deviation $STED_b, b=1$
Level 3: noise classifier	Band-Periodicity $BP_b, b=1,...,B$, Band-Entropy $BE_b, b=1,...,B$

$$SPL = 20\log_{10}(\lambda/\lambda_{ref}) \quad (1)$$

where λ denotes the root mean square value of sound pressure and λ_{ref} the pressure of the lowest sound level that a user of a hearing improvement device can hear. Usually the SPL of quiet environments is