DETECTION OF SIGNALS IN NOISY ENVIRONMENT

Vladimír Sedlák[1], Daniela Ďuračková[1] and Roman Zálusky[1]

[1] *Institute of Electronics and Photonics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava*

E-mail: vladimir.sedlak@stuba.sk

Received 30 April 2012; accepted 03 May 2012.

1. Introduction

The observed signal, in many signal processing applications, can be modelled as a superposition of a finite number of signals embedded in additive noise. Noise is the unwanted energy that interferes with the ability of the receiver to detect the wanted signal. It may enter receiver for example through the antenna along with the desired signal or it may be generated within the receiver (or sensor).

Signal detection deals with the detectability of signals and controlling the criterion that are used for the detection of signals. The task is to find signal that is hides in noise. The relationship between signal and noise is described by Signal-to-Noise Ratio (SNR) of input signal.

SNR in a receiver is the signal power in the receiver divided by the mean noise power of the receiver. All receivers require the signal to exceed the noise by some amount. Usually if the signal power is less than or just equals the noise power it is not detectable. For a signal to be detected, the signal energy plus the noise energy must exceed some threshold.

2. Voice activity detector

Special type of signal detector (VAD) is voice activity detector which works with speech signals. The function of the VAD is to distinguish active speech from non-speech in utterances. It plays an important role in variant speech communication systems, such as speech coding, speech recognition, speech enhancement, and so on. Its accuracy affects their performance. Especially in adverse environments a robust VAD can significantly improve these systems' performance [1].

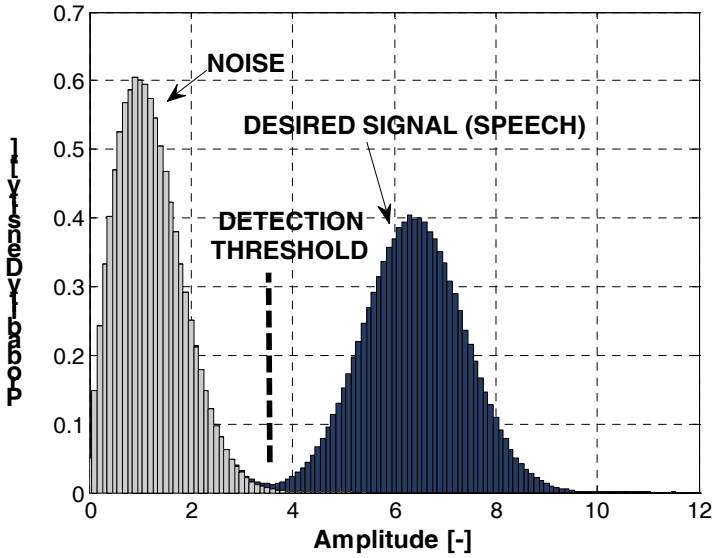

Fig.1: *PDF's of noise and desired signal*

In many areas of speech processing, it is difficult to determine the presence of speech signal, in a given signal. This task can be identified as a statistical hypothesis problem and its purpose is the determination to which category or class a given signal belongs. The detection is based on an observation vector, also called feature vector, which serves as the input to a decision rule that assigns a sample vector to one of given class. The selection of the best feature vector for signal detection and a robust decision rule is a challenging problem that affects the performance of VADs working under noise conditions [2].

Very often are VADs evaluated in terms of ability to discriminate between speech and pause period at different SNR levels (20 dB, 15 dB, 10 dB, 5 dB, 0dB and -5dB). These noisy signals have been recorded at different places (train, car, street, office, etc.)

3. VAD algorithms

Different methods of VAD structure and algorithm have been studied in the past (In the figure 2 is shown the basic block diagram of VAD). Typically, VADs are based on physical differences between a desired signal (speech) and noise. The main VAD designs are based on estimating features such as zero-crossing rate, the periodicity, the energy (Fig. 1 are shown energy differences between noise and desired signal, based on these is chosen the threshold) or the inter-microphone correlation of the signals. Although very complex VADs exist, it is commonly known that very good results are only obtained at high SNRs and when using stationary noise conditions.

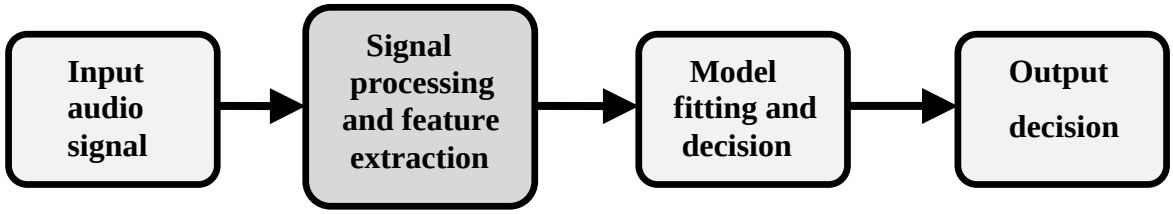

Fig.2: *Block diagram for VAD algorithms*

Zero Crossing Rate-based methods: These methods are based on the computation of the zero-crossing rate (the average number of sign changes in the noisy speech signal *y[k]*). The zero-crossing rate for noise is assumed to be considerably higher than for speech. This assumption is however only accurate at high SNR. At low SNR problems occurs especially in the presence of periodic background noise. The zero-crossing rate can be computed by formula 1.

$$\alpha[k] = \sum_{l=0}^{L-1} \left| sign(sign(y[k-l] - sign[k-l-1])) \right| \qquad (1)$$

Log-energy based methods: Energy-based methods assume that the short-time energy of a speech-and-noise segment is higher than the short-time energy of noise-only segment. By continuously monitoring the signal energy on a frame-by-frame basis, the start and the endpoint of speech can be found when the short-time energy is higher than a threshold value. For this technique is important to the energy of desired speaker was sufficiently larger than the energy of the undesired speakers.

Fourier Transform (DFT)-based methods: These methods are based on discrete Fourier transform (DFT) and they are preferred for good performance and relatively low complexity. DFT-based methods estimate the clean DFT coefficients by applying either a gain function to the noisy DFT coefficients or to the magnitude of the noisy DFT coefficients.

There are more many methods than the above, for example: linear prediction coefficients, Cepstral coefficients, spectral entropy, least-square periodicity measure, wavelet transform coefficients.

4. Applications

VADs are employed in many areas of speech processing. They are very often used within the field of speech communication for achieving high speech coding efficiency and low-bit rate transmission. Another important area is speech enhancement and one of the most popular methods for reducing the effect of background noise is spectral subtraction. In [4] authors present these method is system for speech recognition. The basic idea is estimation of noise spectrum *N(f)* during speech inactive periods and subtracted from the spectrum of the current frame *X(f)* resulting in an estimate of the spectrum *S(f)* of the clean speech (1).

$$|S(f)| = |X(f)| - |N(f)| \tag{2}$$

VADs are also implemented in systems for speech recognition and their qualities strongly influence the performance of these systems. Most of these systems are based on neural networks (NN), hidden Markov models (HMM) or Gaussian mixtures models (GMM) that are trained using a training speech database. The differences between training conditions and real conditions have a deep impact on the accuracy of these systems [3]. In [5] authors presented recognition system for isolated word. VAD is here used for dividing input speech into isolated words. If the VAD determines that it is received the whole word, the system starts the features extraction from this word. These features are then processed by classifier, what is in this case the neural network.

5. Experiments

We used speech signals presented in the [6] for investigation impact of environment for performance of VAD. These signals were obtained from IEEE sentence database and were recorded in a sound-proof booth using Tucker Davis Technologies. These records were produced by three male and three female speakers and were originally sampled at 24 kHz and down-sampled to 8 kHz. The noise signals were taken from the AURORA database (Hirsh et al., 2000) and included the following recordings from different place: car, exhibition hall, restaurant, street, airport, etc. This noise was artificially added to the speech signals using Intermediate Reference System filters at the SNRs of 0 dB, 5 dB, 10 dB and 15 dB. In the figure 3 is depicted input speech signal and output from VAD (energy-based method). The input signal wasn't destroyed by noise and so detection worked very well.

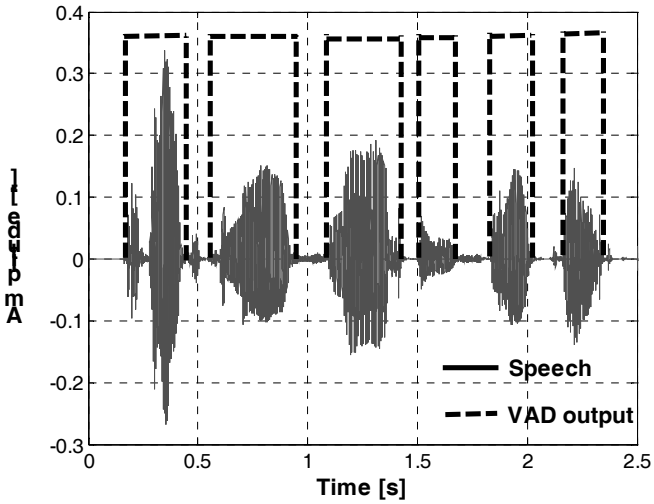

Fig.3: *Input speech and output from VAD*

We evaluated the performance of our VAD (energy-based method) using different types of input signals that were destroyed by different values of the noise. We used the metrics P_{SPEECH} (probability of correctly detecting speech frames) for measure of

performance. This metrics is computed as the ratio of correct speech detections to the total number of hand-labeled speech frames. Achieved results are summarized in table 1.

Tab. 1. *Probability of correctly detecting speech frames*

Environment / SNR [dB]	Car	Street	Exhibition hall	Restaurant
15	96.3%	94.2%	93.2%	95.7%
10	89.2%	87.1%	87.5%	88.1%
5	77%	72.8%	70.9%	74.6%

6. Conclusion

This paper deals with detection of signals in noisy environment. We focused on speech signals but most of mentioned methods can be also applied on other types of signals. Important is right extraction of features from desired signal classifier selection. We have to know character of desired signal as well. Based on simulated results in MATLAB we can see that important factor for right detection of speech is value of additive noise. So it is suitable pre-processing of the input signal using filter (limit the signal spectrum) before the signal is processing by VAD.

Acknowledgement

This work is resulting from the project VEGA 1/0987/12 sponsored by Ministry of Education, Slovak Republic.

References:

[1] D. Ying, Y. Yan, J. Dang, and F. K. Soong: *Voice Activity Detection Based on an Unsupervised Learning Framework*, IEEE Transaction on Audio, Speech and Language Processing, Vol. 19, November 2011.

[2] U. Shrawankar, V. Thakare: *Voice Activity Detector and Noise Trackers for Speech Recognition System in Noisy Environment*, International Journal of Advancements in Computing Technology, Vol. 2, October 2010.

[3] J. Ramírez, J. M. Górriz and J. C. Segura: In: *Robust Speech Recognition and Understanding*, M. Grimm and K. Kroschel (ed.), June, Vienna, Austria, pp. 480 (2007).

[4] J. Han, S. Kim, K. Kim and Y. Yun: *Speech Enhancement Based on Spectral Subtraction for Speech Recognition System*, IEEE International Conference on Consumer Electronics, April 2011.

[5] A. M. Aibinu, M. J. E. Salami, A. R. Najeed, J. F. Azeez and S. M. Ataul: *Evaluating the effect of Voice Activity Detection In Isolated Yoruba Recognition System,* In: International Conference on Mechatronics, May, Kuala Lumpur, Malaysia, (2011).

[6] Y. Hu and P. C. Loizou: *Subjective comparison and evaluation of speech enhancement algorithms*, Speech Communication 49, 2007