Enhancement of speech from noisy background, using single-channel method of spectrogram mapping

Zdenek Smékal, Petr Sysel, Ivan Koula

Department of Telecommunications, Brno University of Technology,
Purkynova 118, 612 00 Brno, Czech Republic,
E-mail: smekal@feec.vutbr.cz

Methods for the enhancement of useful audio signal (speech or musical signal) from interfering background such as low-frequency noise, high-frequency noise, periodic or impulse interference, are of much importance in particular in mobile communications, in voice over the Internet transmission, in various means of transport, etc.

Methods for suppressing undesirable interference in speech signals can be divided into two groups. They are either the single-channel methods (the signal is detected by one microphone) or the multi-channel methods (several microphones or an array of sensors). At present the single-channel methods are given preference in practice since they are considerably simpler as regards the realization of detection elements or the implementation of algorithms. The existing single-channel methods make use of methods for spectral subtraction, adaptive filtering with different modifications of type LMS algorithm, the wavelet transform or digital filter bank, etc. In most cases these methods assume the presence of additive noise, whose properties are close to those of additive noise. A number of firms manufacturing mobile phones and hands-free sets apply the RASTA (RelAtive SpecTrAl) method [1]. This method uses the filtering of variable temporal trajectories of the harmonic components of the Fourier spectrum of speech signal by a low-pass or band-pass digital filter. The type of filter is chosen on the basis of the useful signal-to-noise ratio. The filters suppress the modulation components below 1 Hz and above 16 Hz. Problems arise when noise components are in a frequency range that corresponds to the speech modulation spectrum, and by their magnitude are comparable with the level of speech signal.

In the Department of Telecommunications of Brno University of Technology a novel single-channel method has been developed that makes use also of the short-time spectra of noisy speech signal, similar to the RASTA method. But there is a difference in the way that noise and interference are suppressed. In the time-frequency Fourier representation the regions of speech activity are found out and the interference threshold is selected adaptively. A binary mask is then made, which is used to suppress noise and interference [2, 3]. The method can suppress stationary and non-stationary noise and interference. It is known that speech contains, in unvoiced segments in particular, also noise that must be preserved in order to maintain good intelligibility and satisfactory quality of the reconstructed speech signal. The basic problem of the proposed method consists in determining a threshold that will decide which noise and interference components are undesirable and, vice versa, which have to be preserved since they are part of the speech signal. In the paper, the fundamental principle of the proposed method and ways of determining the adaptive threshold for different types of noise and interference will be described. In conclusion, the ways and problems of implementing the method in type VLIW digital signal processor (TMS 320C6711) will also be discussed.

References:

[1] Hermansky, H., Wan, E.A., Avenando, C.: Speech Enhancement Based On Temporal Processing. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, May 1995, pp.405-408.

[2] Smékal, Z et al. “Non-Linear Methods of Speech Enhancement”. Partial research report on the solution of international project No OC277.002 for the year 2003, Dep. of Telecommunications, Brno University of Technology, 2003, 92 pages.

[3] Smékal, Z et al. “Non-Linear Methods of Speech Enhancement”. Partial research report on the solution of international project No OC277.002 for the year 2004, Dep. of Telecommunications, Brno University of Technology, 2004, 107 pages.