Application of Mel-scale Filter bank for Noise Estimation in Speech Processing

Petr Motlíček
Faculty of Information Technology, Technical University Brno
Phone: +420-5-41149-143, Fax: +420-5-41149-244
E-mail: motlicek@urel.feec.vutbr.cz

Abstract:

The paper addresses the problems of noise estimation algorithm for front-end distributed speech recognition (DSR) system. We propose the method, where Mel-scale filter bank is applied for purpose of spectral smoothing. The selection of noise estimation algorithm is obviously influenced by the following noise suppression system. In our experiments, the noise suppression algorithm has been a-priori given. The memory size limitation have played important role in our noise estimation experiments.

1 Introduction

Speech enhancement algorithms based on spectral decomposition or subtraction, such as Wiener filtering needs accurate estimation of the background noise and very often estimation of signal-to-noise ratio (SNR) for each spectral band independently. This contribution presents some options to estimate the spectral parameters of noise without speech pause detection. However some standard methods that would probably be very successful in our task brought the problems of memory size requirement. Since this limitation is given a-priori, we were obliged to make some modifications.

2 Noise suppression system

The noise suppression algorithm, which is being used in our front-end system, has been derived from standard spectral subtraction and Wiener filtering [2]. The mathematic description of the noise suppression filter is:

$$|H_k[n]|^2 = max\left(\frac{|X_k[n]|^2 - osub|N_k[n]|^2}{|X_k[n]|^2}, \beta\right)^2. \tag{1}$$

The parameters of filter ($osub, \beta$) are different for each frame. Oversubtraction factor $osub$ has to be properly estimated from energy of signal and noise per frame. The spectral floor parameter β prevents the filter components from small values.

3 Estimation of noise spectrum

Often used noise estimation algorithm, proposed in [4], which does not need information about speech/non-speech segments, has been applied in our experiments. The estimation of noise is as well done in power spectra. In temporal domain each spectral subband is filtered by nonlinear estimator that can be perceived as efficient implementation of temporal minima tracking in power spectral domain.

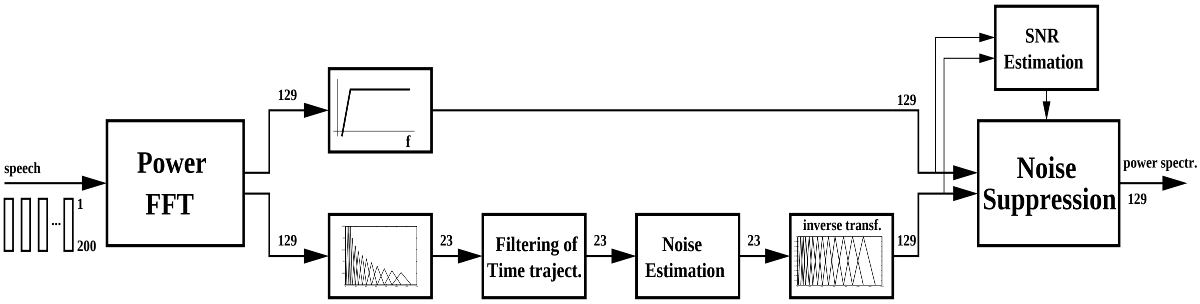

Figure 1: Scheme of noise suppression system with noise estimation applied in Mel-scale filter bank domain.

The previous method is derived from the basic minima tracking algorithm, described in [3]. In this approach noise is estimated as a weighted minimum within a shifting window of appropriate size in temporal domain of power spectra $|X_k[n]|^2$. The shifting window is in practice a memory buffer where p previous values of $|X_k[n]|^2$ are stored. The minima tracking algorithm is being applied independently for each subband of $|X_k[n]|^2$ so that the final buffer size is multiplication of p by number of spectral subbands r.

3.1 Spectral domain processing

For our noise suppression algorithm of DSR system we have experimented with many variations of previously presented noise estimation approaches. The best results obtained on testing databases are mentioned in Tab. 1 (method1, method2). As can be clearly seen from Tab. 1, better results are guaranteed when we apply basic temporal minima tracking algorithm for given noise suppression system.

However the front-end part of DSR system is restricted in many dimensions as for instance memory requirements or algorithm complexity. The minimum size of memory buffer for minima tracking algorithm is 100 (p) x 129 (r) floats which highly exceeds our memory potential.

One of the possibilities to reduce the size of memory buffer is to decrease the number of subbands of input power spectra $|X_k[n]|^2$. In our previous experiments we have observed good behavior of noise estimation algorithms when processing in temporal as well as spectral domain. Therefore we came up with initial transformation of power spectra into critical band energies applying the Mel filter bank. This operation can be perceived as smoothing power spectra in spectral domain. The proposed noise reduction algorithm is shown in Fig. 1. At the beginning the power spectra $|X_k[n]|^2$ is computed using FFT algorithm. Then the input $|X_k[n]|^2$ is split into two branches. In the upper branch (Fig. 1), the DC offset filter is applied and the output is kept as the noisy power spectra for noise reduction algorithm. The DC offset filter is obviously applied in spectral domain of power spectra and is realized as a high pass FIR filter.

In the lower branch of Fig. 1, the input $|X_k[n]|^2$ is smoothed by first order recursive filter in temporal domain [3]. The output power spectra is integrated using Mel filter bank into critical bands. In this transformed spectrum, the noise estimation is done.

Number of subbands of initial power spectra $|X_k[n]|^2$ is 129 $(1 \dots F_{sampling}/2)$. After application of Mel filter bank, 23 critical band energies were obtained. The estimated noise $|N_k[n]|^2$ from Eq. 1 is however expected in power spectral domain (again 129 subbands). Hence we had to apply inverse transformation from 23 critical bands to 129 subbands of power spectra.

In order to keep the same energies in critical bands, standard Mel filter bank for direct transformation was modified, such as:

$$Mfb_{MOD_k}[i] = \frac{Mfb_k[i]}{\sum_{i=1}^{129} Mfb_k[i]}, \quad k \in 1 \dots 23, i \in 1 \dots 129. \tag{2}$$

The $oversub$ constant in Eq. 1 is estimated from SNR ratio of each frame.

4 Testing data

The noise suppression system was tested on three SpeechDat - Car (SDC) databases used for Advanced DSR Front-End Evaluation: Italian SDC [1], Spanish SDC, and Finish SDC.

For each database, three experiments were run. These experiments differ in training conditions. For each of these three conditions, 70% of the files were used for training, 30% for testing. **Well-matched condition (wm):** All the files (close-talk and hands-free microphones) were used. **Medium mis-matched condition (mm):** Only recordings made with the hands-free microphones were used. **Highly mis-matched condition (hm):** For the training only close-talk microphone recordings were used, whereas for testing the hands-free files were taken.

5 Experimental results

The results for all three SDC databases are mentioned in Tab. 1. The results are obtained from DSR front-end system, where another preprocessing blocks, such as voice activity detector, online normalization or another temporal filtering were applied. These blocks have not been touched at all. The output features for speech recognizer were in form of MFCC.

6 Conclusions

The proposed noise estimation system combines the ability of spectral smoothing with minima tracking algorithm in temporal domain of power spectra. This minima tracking noise estimation itself, which guarantee to be very robust for our speech enhancement

$Accuracy$[%]	condition	Italian	Finish	Spanish
baseline	hm	85.01	88.15	88.21
	mm	91.17	86.85	90.67
	wm	96.00	95.48	95.84
method1	hm	88.98	91.98	91.67
	mm	92.61	88.85	92.96
	wm	96.66	96.92	97.22
method2	hm	89.29	92.08	92.66
	mm	93.37	90.56	93.15
	wm	96.82	97.04	96.70
method3	hm	88.61	92.01	91.91
	mm	92.73	88.78	93.53
	wm	96.61	96.97	96.83

Table 1: Speech recognition results for Italian, Finish and Spanish digit databases. *Baseline:* System without noise suppression. *Method 1:* Application of nonlinear filter for noise estimation. *Method 2:* Basic minima tracking in temporal domain. *Method 3:* Noise estimation using minima tracking in Mel-scaled spectrum.

system of front-end DSR does not pass the criteria of memory buffer's size. As described in [3], minimal length of the shifting window is 100 frames. This window can be decomposed into several subwindows (5 in our experiments) without any significant degradation.

The size of used Mel filter bank is 23. Since the noise estimation processing is done in Mel-scaled domain, the size of memory buffer is 23 x 5 floats. However, the results were slightly worse on contrary to the application of basic minima tracking algorithm in 129 spectral subbans of $|X_k[n]|^2$.

References

[1] U. Knoblich. Description and Baseline Results for the Subset of the SpeechDat-Car Italian Database used for ETSI STQ Aurora WI008 Advanced DSR Front-End Evaluation, Alcatel, April 2000.

[2] S. Dupont. Noise Compensation and the Aurora tasks, October 2001.

[3] R. Martin. Spectral Subtraction Based on Minimum Statistics. In *Proc. of EUSIPCO-94*, pp. 1182-1185, Edinburgh, Scotland, U. K., September 1994.

[4] G. Doblinger. Computationally efficient speech enhancement by spectral minima tracking in subbands. In *EUROSPEECH'95*, pp. 1513, Madrid, Spain, September 1995.