Implementation of time-varying modulation filter in speech enhancement system

A.Shadevsky, A.Petrovsky

Computer Engineering Department at the Belarusian State University of Informatics and Radioelectronics, 6, Brovka, Minsk, Belarus
palex@it.org.by;

Abstract
Modern telecommunication technology is present in many areas of everyday life. However for systems carrying out such problems as sound coding, automatic speech recognition, voice identification, etc. is required a high quality signal on an input. This paper describes one-channel neuromorphic motivated speech enhancement system, which explores properties of modulation spectrum of human. It can be used as input signal preprocessor of such systems, with the purpose to increase of its quality.

The new adaptive technique of filtering of spectral envelopes is suggested in this paper. This approach allows more accuracy tune modulation filter, and appropriate to speech enhancement systems. Criteria of a choice of controlled modulation filter parameters are submitted. Also work performances of speech enhancement system are resulted.

1. Introduction
Functioning of a cochlear as the element of an ear can be described at an electric level as work of bank of filters with a high degree of overlapping of passbands. There are a lot of noisy and reverberant speech enhancement methods, which explore properties of modulation spectrum of human speech and work like human ear. The main idea [1] of these methods is splitting speech signal on frequency bands and filtering amplitude envelopes in each band.

According to [2, 3, 4] modulation components of speech basically (more than 95 %), are concentrated in a range from 1 to 16 Hz. It’s caused by quantity of syllables uttering by the person per one second. Thus, the noise varied with frequency out of range, can be removed with a filtration of modulation spectrum. In [4, 5, 6] for this purpose were proposed band-pass filters (1-20 Hz) with different gain-frequency characteristics. Lack of these methods is that pure speech is filtering as well as noisy. In [7-9] were made attempts of environment evaluation by results of which appropriate filter was undertaken. In [7] instead of one modulation filter with the constant gain-frequency characteristics their set was used. The choice of the filter depends on the ratio of energy of a speech component in a band, to the common energy in same band. This approach allows system to choose the modulation filter the most appropriate to conditions of an environment. In work [8, 9], conditions of an environment also were estimated, however instead of a set of filters was used the time-driven modulation filter. In this works were developed pre-processor of speech enhancement for cochlear implant. The goal of ones was extracting speech features in deferent adverse acoustic environment. At installation cochlear implants the person studies to identify speech anew. Therefore the problem of achievement of quality of speech ordinary for communication systems was not put. In this connection the given methods had some lacks which do not allow using them in multimedia systems. First, the central frequency of the filter was fixed. Second, the parameter estimating environment conditions was too much sensitive and reacted to each syllable. It led to fast change of gain-frequency characteristics of modulation filter which constantly was in a condition of transients.

The new adaptive technique of filtering of spectral envelopes is suggested in this paper. This approach allows more accuracy tune modulation filter, and appropriate to speech enhancement systems.

2. Speech processing in modulation domain
Algorithm of speech enhancement [7-9] based on filtration of modulation spectrum with time-varying modulation filter is as follows (see Fig. 1):

1. Speech signal \( x(n/f_s) \) is split into M frequency equal bands using polyphase analysis filter bank.
2. Transformation \( T \) of amplitude envelope \( y_k(nM/f_s) \) is performed using nonlinear static compression in each k-th band \( \tau_k(nM/f_s) \).
3. Filtering the time trajectory of each transformed spectral component \( \tau_k(nM/f_s) \) with time-driven modulation filter \( \pi_k(nM/f_s) \). Simultaneously, modulation filter is recalculated.
4. Filtered speech representation \( y'_k(nM/f_s) \) is transformed \( T^-1 \) back to the linear scale.
5. Speech signal \( \pi(n/f_s) \) is synthesized from filtered magnitude and original phase.
3. Time-driven modulation filter

3.1. Criteria for modulation filter control

As mentioned above, the main part of speech components energy (about 95%) is concentrated in a range from 1 to 16 Hz of modulation domain. In additional, the frequency components higher than 16 Hz have not significant influence on the quality and intelligibility of speech signal. Besides, increasing of low cutoff frequency of modulation filter higher than 1 Hz [3] leads to degradation of speech quality and intelligibility. According to experimental data changing of low cutoff frequency from 0 to 1 Hz exert lower influence on speech quality on system output than changing of stopband attenuation. Because above-stated, we use bandpass filter with fixed high cutoff frequency 16 Hz (see Fig. 2) for filtration in modulation domain. The low cutoff frequency also is fixed 1 Hz, but it stopband attenuation is tunable by $a_k$.

![Figure 2. Criteria for modulation filter control](image)

Such approach allows to tune modulation filter to changing acoustical environment and to attenuate main part of noisy components. However, there is such drawback as residual noise, because the results of such approach to be comparable with the conventional spectral subtraction [2]. For disposal of this lack the second parameter $b_k$ is used. Second parameter $b_k$ is used for disposal of this lack. It manages of passband attenuation. Its value is determined by present of speech components in the frequency band. Thus, the parameter $b_k$ is the peculiar voice activity detector.

3.2. Filter parameters computation

According to Fig. 3, transformed by cubic root amplitude envelope $y'_k(nM/f_s)$ (Fig. 4b) on the input of MF parameters computation block is filtered by band-pass filter (1-16 Hz) for estimation speech component. It is assumed that signal $y'_k(nM/f_s)$ contains only speech components.

The filtration of amplitude envelope by HPF deletes low frequencies of speech component. As consequence average value of a filtered signal aspires to zero (Fig. 4c). Besides in it negative values have appeared. Therefore restoration of amplitude speech components under the formula is carried out (Fig. 4d):

$$s'_i = y'_i + \sigma_i$$  \hspace{1cm} (1)

where $\sigma$ - standard deviation of filtered envelope $s'_i(nM/f_s)$ (Fig. 4c), calculated for each sample into frame with length $L$. Length $L$ gets out according to that a maximum of speech components energy are concentrated in area 3-5 Hz. Therefore frame length corresponds to the period of frequency of 4 Hz:

$$L = f_s / 4M$$  \hspace{1cm} (2)

Now we know amplitude envelope of speech component and can detect amplitude envelope for noisy components (Fig. 4f) as:

$$n'_i(nM/f_s) = y'_i(nM/f_s) - s'_i(nM/f_s)$$  \hspace{1cm} (3)
The computation of the ModulationFilter (MF) parameters is performed through the following steps:

1. **Modulation Filter**

   - **Delay**
   - **Computation of MF parameters**

   The STFT operation of signals \( n'(nM/f) \) and \( \pi_t(nM/f) \) is carried out for the computation of modulation filter parameters in the \( k \)-th band. The parameter \( a_k \) is computed as:

   \[
   a_k = \frac{Y_k(0) - N_k(0)}{Y_k(0)}
   \]  

   where \( Y(0), N(0) \) - zero frequency amplitude of current frame for signals \( \pi_t(nM/f) \) and \( n'(nM/f) \) respectively. Thus, the parameter \( a_k \) defines in how many times it is necessary to weaken a signal in a stopband (0-1 Hz).

2. **Passband filter**

   The noise level in a passband 1-16 Hz is estimated on the basis of noise level in a range over 16 Hz. The parameter \( b_k \) is calculated under the formula:

   \[
   b_k = \frac{mean(Y,(1-16)) - mean(N,(17-32))}{mean(Y,(1-16))}
   \]  

   where \( mean(Y,(1-16)) \) - mean value of spectral amplitude in current frame for signal \( \pi_t(nM/f) \) for 1-16 Hz frequency components, \( mean(N,(17-32)) \) - mean value of spectral amplitude in current frame for signal \( n'(nM/f) \) for 17-32 Hz frequency components. Thus, the parameter \( b_k \) defines in how many times it is necessary to weaken a signal in a passband (1-16 Hz).

---

**Figure 3. Computation block of MF parameters**

**Figure 4. Filter parameters computation:**
- a) speech signal with additive dynamic noise; b) amplitude envelope \( \pi_t(nM/f) \) (band 448-480 Hz); c) amplitude envelope filtered by bandpass filter \( \pi'_t(nM/f) \); d) amplitude envelope after restoration \( n'(nM/f) \); e) standard deviation of filtered envelope \( \sigma \); f) amplitude envelope of noisy components \( n'(nM/f) \); g) \( b_k \) parameter; h) \( a_k \) parameter.
Figure 5. – Clean speech signal and its spectrogram.

Figure 6. – Noised speech signal (white noise 0dB, reverberation, tone 450 Hz) and its spectrogram.

Figure 7. – Processed speech signal and its spectrogram.
Amplitude-frequency response is computed on the basis of the designed parameters. Then spectral envelope \( r_v(nM/f_v) \) is filtered by the modulation filter in frequency domain.

Frequency response of the modulation filter is changed in each \( k \)-th band in time for each frames of signal \( r_v(nM/f_v) \). Modulation filter parameters \( a_k \) (Fig. 4h) and \( b_k \) (Fig. 4g) are shown in Fig. 4. Apparently from figure, parameter \( a_k \) is decreased at occurrence of noise, thus, attenuation degree of the modulation filter in a band of 0-1 Hz is increased. And parameter \( a_k \) is increased when noise level is small. Parameter \( b_k \) is increased at occurrence of speech components in the channel and is decreased in pauses.

Time constant of given algorithm determine value of delay in modulation filter channel.

4. Experimental results

Complex noised speech signal with sampling frequency 8 kHz has been used to demonstrate offered method work results. It was modulated with additive white noise (SNR = 0 dB), reverberation and tone 450 Hz (SNR = -5 dB) (see Fig. 6). The source signal and its spectrogram are represented on Fig 5. The processed signal and its spectrogram are shown on Fig. 7. For the spectrogram easily can bee seen phonemic structure of speech, and non speech components are significant attenuated. SNR of processed signal in noised region is 24 dB and in tone region is 65 dB. Reverberation was reduced.

Also experiments with convolutional noise and noised speech signals with SNR below 0 dB were carried out. In those experiments increasing speech quality and intelligibility was marked too.

It is necessary to note, that signals with SNR <0 dB have lower intelligibility and higher quality after processing than noised speech signal.

5. Conclusion

Proposed algorithm of enhancement of noisy speech explores properties of modulation spectrum of human speech and adapts to changing acoustical environment. It eliminates slow time-varying noises. It doesn’t require reference channel with noise signal. Proposed approach allows more accuracy tune modulation filter, and appropriate to speech enhancement systems. Method efficiently deals with additive and convolution noises, reverberation.

6. References