BROADBAND NOISE CANCELLATION SYSTEMS: NEW APPROACH TO WORKING PERFORMANCE OPTIMIZATION

Serguei Koval, Mikhail Stolbov, Mikhail Khitrov
Speech Technology Center, St.-Petersburg, Russia
Tel.: +7(812) 3258848; Fax: +7(812) 3279297; E-mail: koval@speechpro.com

ABSTRACT

The report deals with speech enhancement in a noisy environment. A new, direct shaping parametric formulation of generalised adaptive spectral subtraction algorithm for non-stationary interference is considered. The idea of the authors is not to select but to form desirable envelope of spectral estimator function in SNR\*Gain-function area. Some principles of parameters choice are proposed, based not only on aural estimations of signal quality but also on graphical analysis of target test signals. The field examinations showed good working performance of the proposed approach for noisy speech signal enhancement. The accuracy of various speech recognisers with the described pre-processing was significantly improved for noisy input speech. The paper develops approach originally reported in [5].

INTRODUCTION

Speech recognition, coding, transmission and processing systems operating in adverse environments, such as moving car or noisy office, street, etc. have to deal with a variety of ambient noises. Noise affects speech recognition or processing, and the result is a rapid deterioration in working performance of practical systems with decreasing signal-to-noise ratio (SNR).

The majority of real interference types are additive non-stationary broadband noise. To decide a task of speech signal clearing from an additive broadband interference the method of spectral subtraction (SS) is the most effective one, because its approach corresponds to the nature of such a speech corruption in the best degree. From the time of the classical formulation of SS-based speech enhancement technology [1] there were published its numerous modifications and improvements. But up to now this method has “a bad reputation”: usage of this approach in field applications is limited by low subjective output speech quality. All researchers noticed that although the original SS method is capable of noise suppression it often results in excessive residual noise including unpleasant "musical tone artefacts". Useful speech signal after SS processing has got some "smearing" and corruption. As an example we can quote [11]: “The processing distortions in spectral subtraction is a fundamental limitation of this method and is the main cause for the relatively poor performance of spectral subtraction”. Even the very last works have not so big achievements to overcome this SS disadvantage. Today’s speech enhancement systems using SS are practically acceptable for very limited types of noises and low levels of background noise or they require an obligatory manual adjustment of SS control parameters “by ear". It caused as a matter of fact the absence of "general purpose" automatic effective means and tools for broadband noise removal in voice communication, automatic speech recognition and speech processing for the hearing impaired.

In the report an original insight into SS methods for speech enhancement is proposed. The report develops the authors’ ideas published originally in [5]. Proposed ASS was tested for many field applications and has excellent performance.

BACKGROUND OF THE METHOD

There is an input signal x(t), consisting of a speech signal s(t) and interference n(t), which is an input signal of noise cancellation procedure. It is supposed that a signal and interference are quasi stationary, i.e. their statistical characteristics are changed in time not too fast. The expression of the optimum noise reduction filter e.g. in the form of classical spectral subtraction can be given by equation of spectral estimator [1]:

\[
\hat{\Omega}_s(f, t) = \hat{\Omega}_x(f, t) - \hat{\Omega}_n(f, t),
\]

where f - denotes frequency, t - time, \(\hat{\Omega}_x(f, t)\), \(\hat{\Omega}_s(f, t)\), \(\hat{\Omega}_n(f, t)\) - power spectrum density estimates of an input signal, speech and interference correspondingly.

The classical ASS filter is constructed as follows. The spectrum estimate of non-stationary interference \(\hat{\Omega}_n(f, t)\) is calculated in pauses of speech. Transfer function g(f,t) of the noise reduction filter can be written as:

\[
g(f, t) = \frac{\hat{\Omega}_s(f, t)}{\hat{\Omega}_s(f, t) + \hat{\Omega}_n(f, t)} = \frac{\hat{\Omega}_s(f, t)}{1 + \text{SNR}(f, t)}
\]

where \text{SNR}(f, t) = \frac{\hat{\Omega}_s(f, t)}{\hat{\Omega}_n(f, t)} - current signal-to-noise ratio (SNR) in the given frequency band.

One can interpret a filter ASS in a following way. For each time moment in spectral band where the noise goes over a useful signal (SNR(f, t) > 1), the filter does not pass a signal, i.e. g(f, t) → 0. There, where the signal is more than the interference, (SNR(f, t) > 1) the filter transfer function g(f, t) → 1 and does not change the signal. All known to authors spectral estimators have in SNR area similar behaviour and differ in the specific envelope form.

Nowadays three sets of methods are proposed: classical based on signal statistics, parametrical which allow to tune noise cancellation parameters and psychoacoustical which adjust processing parameters automatically according human ear perceptual properties.

Based on various models of signals and criteria of an optimality the various expressions for gain function and corresponding spectral estimator of ASS filter were proposed. In the Table 1 one can find some of them: Many previously used methods typically do not provide with sufficient noise reduction or do not keep properly safe essential speech features. The most substantial subjective dimensionalities for quality evaluation of the signal after SS filtering are the level of the residual noise, perceptual level of “music “noise”, distortion degree of useful speech signal (typically speech after ASS may be some “smeared” and “bitten”).

TABLE 1. Various Existing algorithms and Related Noise
### Suppression Functions

<table>
<thead>
<tr>
<th>SS methods</th>
<th>Formula for noise suppression factor $g(f,t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical algorithms</td>
<td>$\frac{\sqrt{\text{SNR}(f,t)}}{1 + \frac{1}{\text{SNR}(f,t)}}$</td>
</tr>
<tr>
<td>Power subtraction [2]</td>
<td>$\sqrt{\text{SNR}(f,t)}$</td>
</tr>
<tr>
<td>Maximum Likelihood [2]</td>
<td>$\frac{1}{2} \left[ 1 + \frac{\text{SNR}(f,t)}{1 + \text{SNR}(f,t)} \right] \Phi_f(x)$</td>
</tr>
<tr>
<td>Soft decision maximum likelihood [2]</td>
<td>$\frac{1}{2} \left[ 1 + \frac{\text{SNR}(f,t)}{1 + \text{SNR}(f,t)} \right] \Phi_f(x)$</td>
</tr>
<tr>
<td>Wiener filtering [3]</td>
<td>$\frac{\text{SNR}(f,t)}{1 + \text{SNR}(f,t)}$</td>
</tr>
</tbody>
</table>

#### Parametric algorithms

<table>
<thead>
<tr>
<th>Parametric Formulation [4]</th>
<th>$\frac{a \Phi_f(f,t) + b \Phi_f(f,t)}{\Phi_f(f,t)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct shaping parametric formulation [5]</td>
<td>$\frac{1 + c \left( \text{SNR}(f,t) - \text{SNRo} \right)}{2 + c \left( \text{SNR}(f,t) - \text{SNRo} \right)}$</td>
</tr>
</tbody>
</table>

#### Psychoacoustic algorithms

<table>
<thead>
<tr>
<th>Psychoacoustic [6]</th>
<th>$\frac{\Phi_f(f,t)}{A(f,T) + \Phi_f(f,t)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on masking properties of the human auditory system [7]</td>
<td>$\frac{1 - \Phi_f(f,t)}{\Phi_f(f,t)}$</td>
</tr>
</tbody>
</table>

One of the effects caused by the ASS method is the appearance in a filtered signal specific “musical” noise – a number of “musical tone artefacts”, due to the presence of random spectral peaks of the residual noise. This residual noise spectral jitter may be contrasted and unmasked after processing and is so more audible. The standard way to get required ASS filter performance was to choose some kind of spectral estimator and test the working features for given estimator increasing efficiency of noise/speech decision procedure for processed signal or using some a priori information about SNR. These points are quite significant, but here we analyse the third way - to form the optimal spectral estimator for the three above mentioned criteria. For simplicity we explain this approach describing the forming of optimal ASS filter gain function in SNR area.

The basic requirements to ASS filter are the following:
1. ASS filter should adjust itself to different SNR, signal and noise levels and speed of changing in every spectral band.
2. It should have simple intuitive obvious functions of control for residual noise level, useful signal safety and “music” noise level in output signal.

We presuppose that besides automatic filter adjustment there should be simple means of control over noise reduction degree and adjustment rate.

Not less important to have obvious criteria to tune the method properly to specific conditions and tasks.

The most important factors for practical realization of the algorithm are the following.
1. Specific kind of filter function (algorithm).
2. Adaptation in time area.
3. Speech/pause criteria (Vocal Activity Detector, VAD).

### METHOD

The purpose of the authors was the development of an ASS filter self-adjusting to conditions of specific acoustic situation, in particular, to various SNR. The offered decision includes: 1- models of transfer function of the filter; 2- procedure of recognition of pauses in speech, 3- updating the information about interference. In this report we discuss the first point.

We decided not to use the filter gain function in any statistically based form [3-7,10-15], but to propose similar, rather “free” parametric expression in order to form the best envelope experimentally by suitable selection of parameters values.

The model of the filter has adjustable parameters: contrast $c$, target SNR - SNRo, minimal spectral gain floor (depth) – Gm or masking threshold, $\alpha$ - shaping factor (typically from 0.5 to 4):

$$g(f,t) = \begin{cases} 1 + c^* \left( \text{SNR}(f,t) - \text{SNRo} \right) \\ 2 + c^* \left( \text{SNR}(f,t) - \text{SNRo} \right) \end{cases}$$

where $c$ is contrast, $\text{SNRo}$ is target SNR, $\alpha$ is shaping factor.

The meaning of the offered model is quite simple. Parameter SNRo allows to control SNR level for a point at which $g(f,t) = 0.5$. Changing SNRo we can control the place of this point in the working curve according to the specific type of noise and the choosen degree of noise reduction.

Some smoothing of transfer function in temporary $(E_t(\cdot))$ and frequency $(E_r(\cdot))$ areas was used as additional means of residual and music noise reduction.

$$G(f,t) \text{ smoothed} = E_r(E_t(g(f,t)))$$

Charts illustrating the dependance of working curve on control parameters are given in Figures 1, 2.
REALISATION AND OPTIMIZATION

The incoming noisy speech signal is frame-by-frame windowed by a half-overlapped Hann window and then spectrally decomposed by the N-points fast Fourier transform (FFT). The spectral amplitude of the noise-suppressed speech for every spectral band is calculated using (3, 4, 5). The spectral amplitude is then combined with the original phase. Inverse FFT and by overlapping and adding two consecutive time frames. Non-optimised software realisation occupies 25 % of a calculating power of Pentium 100 MHz.

Parametric expression of $g(f, t)$ (3,4) as a function of contrast, target SNR, spectral floor and $\alpha$ provides with the possibility to imitate in essential features any before known spectral estimators or to construct in a space $G(f, t)$*SNR the required curve. We can choose the best spectral estimator performance varying the parameters.

By changing of target SNRo it is possible to establish any desirable level (up to zero amplitude) of the residual noise in pauses for quite wide range of SNR (e.g. for white noise as interference at least for SNR of speech components from +60 dB to −14 dB). Every time ASS corresponds to a certain degree of speech components corruption for every SNR in signal. For example if the difference between two parts of the original word was 1 dB of power, then after ASS processing the corresponded difference of the same components may be near 1 dB for relatively low levels of background noise (> +10 dB SNR) and significantly more for the higher relative levels of noise (e.g. for SNR <−4 dB it may be 6-12 dB for typical ASS variants). To our mind it is one of the main reasons of the poor speech quality processed by ASS. Really every word usually has sounds with difference in power. E.g. the word “Six” may have typically 24 dB between [i] and [ks] sounds. So for different sounds and parts of the sounds of an utterance by fixed level of background noise may have quite big difference in SNR, so they would get some big additional difference in power after processing. This phenomenon is manifesting in particular for onsets and releases of separate sounds, providing with the effects of speech “smearing” and “nibbling”. Using constructed test speech signals as a mixture of fixed noise and the number of speech components sounds with the consecutively changing amplitudes we have the line of sounds with different SNR. Then it is possible only graphically, watching in the interesting to us area the SNR difference for those test signals before and after processing to fit the best trade-off of residual noise level against speech distortion degree varying the control parameters in (3,4,5).

The advantage of the optimised spectral estimator in comparison with for example standard ASS filter [4,5] we can describe in the following way. For the same fixed level of the residual noise (e.g. 12 dB less than original noise level) the speech components in processed signal have nearly the same level for both methods for a priory SNR > 10 dB. For the a priory SNR =10 dB standard method has an additional decrease of speech components in 1 dB in comparison to our ASS. Then for less values of a priory SNR the additional decrease of speech components by standard ASS is growing up to 8 dB for SNR=0 dB. For this test standard ASS has non-working performance for speech components with SNR > 2dB, our ASS is working up to SNR - 12 dB. And such an advantage of a new ASS has place for any SNR and any types of noise.

These considerations allow to tune ASS parameters for test signals automatically choosing the best speech components safety with a fixed level of residual noise. The main SNR area of difference between between ASS variants lies in +6 upto −12 dB.

The music noise may be eliminated by increasing the spectral and time intervals of averaging in (5) and $Gm$, that will cause the increase of the residual noise level and therefore the loss in SNR gain. For ASR the presence of music noise has a small effect upon success rate.

RESULTS AND THEIR DISCUSSION

We have investigated the offered algorithm in a test with changing SNR and determine influence of control parameters over ratio of noise reduction and speech reduction.

The opportunities for suppression of residual and music noise with the help of various parameters are investigated. The following influence of parameters of the filter on efficiency of suppression of noise is revealed:

1. With increase of values $\chi$ and SNRo in (3) residual noise decreases in level and turns into "bells", then “bells” become less and less audible, speech losses its quality, further growth of these parameters results in suppression of
both noise and useful signal. It is established, that the developed algorithm provides in field conditions suppression of multi-tonal narrow-band interference to 25 - 70 dB and broadband noise to 10 - 40 dB with acceptable speech quality.

2. The reduction of Gm results in masking of existing music-noise. The subjective optimum range of suppression lies between 10 and 40 dB and depends on SNR in input signal, noise type and application task.

4. The increase of number of spectral bands of the filter results in change of sounding of musical noise: in a line: “scrubbing” - “bells” - “organ”. For many types of real noises the maximal number of filter bands provides the best results, but the big duration of analysis window gives some echo effect. The optimum size of spectrum analysis window depends on the a priori SNR, sampling rate and amounts to 128-4096 components.

5. Smoothing transfer function in frequency areas is an effective mean of music noise suppression.

6. The manual optimisation of ASS filter parameters is useful but does not give very substantial effects in majority of applications and for wide range of SNR and noise types. Of course the individual long-term renovation of old music records by sound artists demands some specialised version of proposed ASS.

7. The objective and subjective criterions for evaluation of residual speech after ASS may have quite noticeable difference in specific application field.

Described here noise cancelling ASS were incorporated into the sound processing Workstation “IKAR” and stand-alone speech enhancement device “Cinderella-97” produced by STC. The method has working name “ClearVoice”. These tools are widely used in law enforcement for practical noisy speech record processing and have shown the real effectiveness in field conditions.

The processed speech signal is recognised as much more useful but does not give very substantial effects in majority of applications and for wide range of SNR and noise types. Of course the individual long-term renovation of old music records by sound artists demands some specialised version of proposed ASS.

7. The objective and subjective criterions for evaluation of residual speech after ASS may have quite noticeable difference in specific application field.

Described here noise cancelling ASS were incorporated into the sound processing Workstation “IKAR” and stand-alone speech enhancement device “Cinderella-97” produced by STC. The method has working name “ClearVoice”. These tools are widely used in law enforcement for practical noisy speech record processing and have shown the real effectiveness in field conditions.

The processed speech signal is recognised as much more useful but does not give very substantial effects in majority of applications and for wide range of SNR and noise types. Of course the individual long-term renovation of old music records by sound artists demands some specialised version of proposed ASS.

CONCLUSION

The new method of an ASS filter realisation is proposed. Numerous tests in field applications showed good working performance of the described method of speech enhancement for additive non-stationary broadband noises. New ASS method features:

• Essential improvement of subjective speech quality and intelligibility for signals with SNR > –12 dB;
• Dramatic increase of accuracy of speech recognition after ASS application for noisy speech signal; no accuracy lowering for filtered clear speech;
• Improvement of noisy speech intelligibility for vocoders and speech coders;
• Automatic detection of the characteristics of speech and interference;
• Suppression of tone to 25 - 70 dB, broadband noise to 10 - 40 dB depending on SNR and application task;
• Opportunity of fine adjustment for specific conditions of noise environment and application task.

REFERENCES