PARAMETRIC MULTI-BAND AUTOMATIC GAIN CONTROL FOR NOISY SPEECH ENHANCEMENT

M.Stolbov, S.Koval, M.Khitrov

Speech Technology Center
St.-Petersburg, Russia
info@speechpro.com, http://www.speechpro.com

Abstract

This report is devoted to a new approach to wide band non-stationary noise reduction and corrupted speech signal enhancement. The objective is to provide processed speech intelligibility and quality while maintaining computation simplicity. We present a new (non-subtractive) noise suppression method called multi-band Automatic Gain Control (AGC). The proposed method is based on the introduction of a non-subtractive noise suppression model and multi band filter gain control. This model provides less residual noise and better speech quality over the Spectral Subtraction Method (SSM). Modification of multi-band AGC gain function allows easy introduce new useful feature called Spectral Contrasting. The report contains the discussion of AGC control parameters values. Experiments show that the proposed algorithms are effective in non-stationary noisy background for Signal-to-Noise Ratio (SNR) up to –6dB.

Keywords: noise reduction, non-stationary environments, spectral subtraction, speech signal enhancement

1.Introduction

The objective of this article is to improve noise suppression performance and provide speech intelligibility and subjective quality (low residual noise and preserving weak speech components), while maintaining its computation simplicity.
SSM is well known and is a very important noise suppression method for many practical tasks. A large number of SSM algorithms are used for different purposes (speech recognition, speech enhancement, telephony, hearing aids, music restoration etc). The main disadvantage in using conventional SSM is the resulting "musical tones" in output signal.

This disadvantage is usually overcome in two ways:
1. Parametric formulations of SSM [1-6]
2. Adaptive tuning of SSM parameters (adaptive averaging of gain function etc)[1,3,5].

This paper addresses parametric formulation of signal enhancement methodology.
Parametric formulation has the purpose of controlling shape of gain function. The choice of parameter values allows tuning those methods to various conditions. Classical algorithms are usually limited by usage of fixed parameters, which are optimal for not all speech and noise conditions. Further more non-stationary noises demand different parameter values.

Another disadvantage of SSM is the difficulty reducing residual noise while maintaining unsuppressed weak speech components, this property follows from shape of SSM gain function (subtraction function). The cut of noise produces the cut of weak speech.

Some alternative (non-subtractive) gain function models were proposed to attenuate noise while preserving speech. For example multi-band Compressor for hearing aids and Adaptive Gaussian Attenuation (AGA) are known [6,7]. In this report we propose an alternative algorithm, ground the choice of control parameters and investigate its properties. Our method is a further development of a parametric non-subtractive methods and is based on our previous ideas [4,5]. The proposed parametric formulation describes the original method and several its modifications.

2. Multi-band automatic gain control

In the spectrum domain the noisy speech is distorted as follows:

$$X(f,t) = S(f,t) + N(f,t),$$

where $S(f, t)$ is power or magnitude energy at time t and frequency band f of the original speech signal, $N(f, t)$ is the noise energy, $X(f, t)$ is the energy of the corrupted signal. We use the magnitude energy.
The SSM gain function Gssm(f) is based on noise spectrum subtraction[1-3]:

$$Gssm(f) = \frac{X(f)^K - b\,N(f)^K}{X(f)^K}, \quad \text{if } Gssm(f) < Go \text{ than } Gssm(f) = Go,$$

where Go is suppression depth (noise floor, flooring factor), $b>=1$ – oversubtraction factor (suppression strength or suppression threshold), $K=1,2.$ So SSM is controlled with two parameters.

We define the multi-band gain function of AGC1: $G1(f)$ by a similar way as the frequency dependent one band dynamic filter gain function:

$$G1(f) = Go\ X(f)/N(f), \quad \text{If } G1(f) > 1 \text{ than } G1(f) = 1$$

This filter uses 1 parameter – the attenuation coefficient Go.

If $X(f)$ is close to noise spectrum component $N(f)$ then $G1(f) \approx Go$. In such a case signal components (and whole signal) would be suppressed in *Go* times. Hence parameter *Go* controls suppression depth as in SSM.

Suppression depth will decrease when signal components increase. If $X(f)$ achieves level $N(f)/Go$ the corresponding component would be unsuppressed. Hence parameter *Go* defines the Input to Noise Ratio INRo = $1/Go$ over that signal components would not be changed. The range of gain function from suppression to non-suppression region is the sharpness or suppression strength of the filter. Easy to see Go is smaller the sharpness is smaller. Hence Go controls at once two properties of AGC1: suppression depth of noise and sharpness.

It's often useful to have lower suppression depth, and sharper working characteristic. We can get it in the follows manner. For any INRo>1 we can find the power K>1 that

$$Go = (1/INRo)^K$$

To have noise suppression depth *Go* and fixed noise suppression limit INRo, we use following 1-parametr filter:

$$G1(f) = Go * [X(f) / N(f)]^K.$$

For example if the suppression depth is -20 dB $(Go = 0.1)$ and INRo is close to *3*, characteristic may be achieved using AGC1 filter with power $K=2$, which is convenient for practical realization of the algorithm.

The advantage of one-parametric AGC1 is its simplicity and and the possibilities to work efficiently under different noise conditions. However the disadvantage is the correlation of suppression depth and sharpness. We can't enlarge the suppression depth without strong speech component suppression. That's why the working suppression range is limited about –12..-18dB.

To overcome these limitations we propose 2-parametric AGC2. We constructed two-parametric AGC2 filters with the following gain function:

$$G2(f) = C \left[\frac{X(f)^K - b*N(f)^K}{N(f)^K}\right] = C \left[\frac{X(f)^K}{N(f)^K} - b\right],$$

If $G2(f) < Go$ *than* $G2(f) = Go$, *if* $G2(f) > 1$ *then* $G2(f) = 1$,

where *C* is *contrast*.

This parametric AGC function allows the choice of independently suppressing depth, suppressing strength and sharpness. We can tune this function for specific tasks.

To construct gain function similar to SSM $(b=1, K=1)$ we use for AGC2 the follows parameters: $b=1$, $K=1$, $C=0.5$.

3. AGC AND SSM COMPARISON

The family of signal processing gain functions $G(f)$ is shown in Table 1.

Table 1. Gain function formulas for SSM, AGA and AGC.

	LINE	QUADRATIC
SSM	$1 - \frac{bN(f)}{X(f)}$	$1 - \frac{bN(f)^2}{X(f)^2}$
AGA	$\frac{1}{1 + C*\exp-\left[\frac{X[f] - bN(f)}{\sqrt{\sigma(f)}}\right]}$	
AGC1	$Go\frac{X(f)}{N(f)}$	$Go\frac{X(f)^2}{N(f)^2}$
AGC2	$C*\left[\frac{X(f)}{N(f)} - b\right]$	$C*\left[\frac{X(f)^2}{N(f)^2} - b\right]$

Gain functions vs. input-to-noise-ratio INR are shown on Fig.1 and Fig.2.

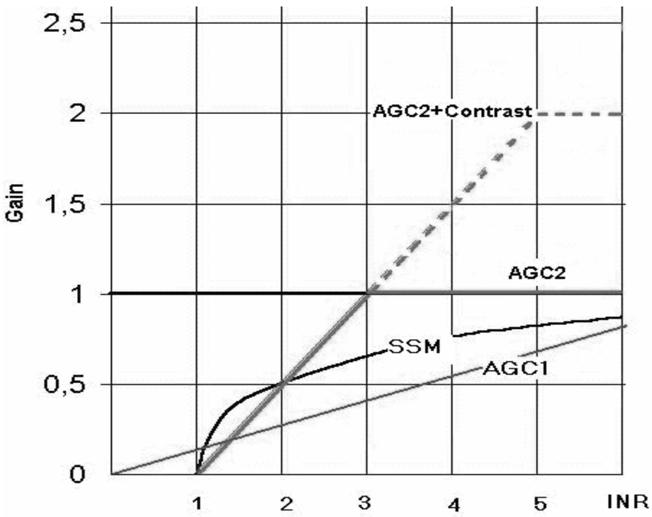

Figure 1. Filter gain functions for different noise reduction methods vs. INR *(b=1, C=0.5, K=1, Go=0.125)*

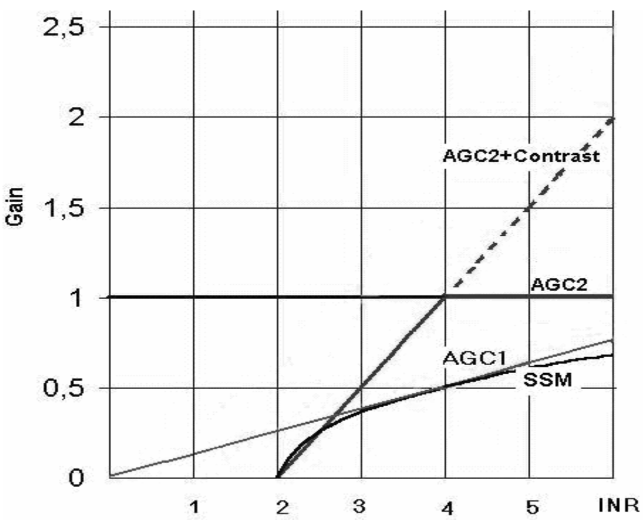

Figure 2. Gain functions vs. INR *(b=2, C=0.5, K=1, Go=0.125)*

Let's compare the SSM (b=1), AGC1 and AGC2 (C=0.5, b=1) methods. We examine output signal distortions – residual noise and speech oversuppression. The sharpness of the gain function leads to sensitivity as to random variations of instant noise amplitude as to bias of average noise estimate.

Residual noise is produced by unequal suppression of noise components. The variations of input noise components produce contrasting variations of output noise components which are so called musical noise.

Over suppression of weak speech components is produced if the average noise component estimate is larger than its real amplitude, which may be caused either by fitting errors or by non-stationary noise variations.

From gain functions formulas and figures 1,2 follows: AGC1 has the least residual noise and the least over-suppression and SSM has the largest ones.

4. Signal spectral contrasting.

AGC gain functions allow intelligibility by amplifying the weak speech components (SNR(f) <=12 dB). We name additional signal spectral components amplification in the regions where signal exceeds background as spectral contrasting. Spectral contrasting may be realized by maximum gain function rate enlargement (under usual $G\_max=1$):

If $G(f) > G\_max$ *then* $G(f) = G\_max$

Where $G\_max > 1$ is amplification bound for signal spectral components.

Gain functions amplitude vs. INR for AGC2 with contrasting are shown on Fig.1 and Fig.2.

5. Experimental results

We compared AGC and SSM efficiency in low SNR range. The useful signal average power safety after filtering was chosen as criteria of comparison.

As a demonstration test signal we used the mixture of the white noise and the signal of the same word "One" decreasing in energy by step 1 dB [4]. On the figure SNR is changing from the left (+2dB) to the right (-8 dB). The power of input mixture (noise + speech signal), original clear speech signal, SSM output signal (Go=-18 dB, b=1) and AGC2 output signal (Go=-18 dB, C=0.5, b=1, G_max= 6 dB) vs. input SNR are shown on Fig.3. Table 2 shows the experimental powers as gain function vs. INR for different noise levels for SNR from +14dB to – 6 dB.

The experimental results showed good accordance behaviour of average output processed signal power with single band gain function. Contrasted AGC2 *(contrast* = 6dB) produces larger power of speech than SSM up to –5 dB. Both methods produce useful output signal power comparable with original clear speech power up to SNR = -6 dB. Subjective listening tests shows AGC1and AGC2 with $Go= -12..-18$ *dB* produces the most intelligible and natural filtered signal for a wide number of real acoustical and channel noise types**.**

6. Conclusions

This paper presents a new approach to non-stationary wideband noise reduction for cases when it is difficult to get correct noise spectrum estimate. A new spectral AGC algorithm and its modifications are proposed. AGC algorithm has follows properties:

- Its noise reduction efficiency is like SSM with SNR up to –6 dB.
- It can be used for nose corrupted speech signals as well as for other signals too.
- AGC is suited for full-automated relatively universal signal processing.
- Computation simplicity.
- It produces subjectively good output sound for hearing purposes (no artifacts after noise suppression).

The model and algorithms were successfully used for both restoration and enhancement of poor quality speech records made in noisy environment such as street, office, machinery halls or through long distance communication channels. Implementation of the described algorithm to the front-end sound processing in hands-free car speech recognizers provide reduction of recognition errors up to 2-3 times for noisy environments of moving cars at high speed and open windows.

Test software and examples are on site http://www.speechpro.com.

7. References

[1] Virag N., *Single channel speech enhancement based on masking properties of the human auditory system*, IEEE Tr. on SAP, Vol. 7, No.2, 1999, pp.126-137.

[2] Sim B.L. et. al., *A parametric formulation of the generalized spectral subtraction method*, IEEE Tr. on SAP, Vol. 6, No.4, 1998, pp.328-336.

[3] Linhard K., Haulick T., *Noise subtraction with parametric recursive gain curves,* Proc. EUROSPEECH"99, 1999,Vol. 6, pp.2611-2614.

[4] Koval, S., Stolbov M., Khitrov M., *Broadband Noise Cancellation Systems: New Approach to Working Performance Optimization,* Proc. EUROSPEECH"99, 1999,Vol.6, pp. 2615-2628.

[5] Koval, S., Stolbov M., Khitrov M., *A new approach to generalised spectral subtraction method as pre-processing technique for robust speech recognizers,* Proc. Int.conf. ROBUST 99, Tampere, Finland, 1999, pp. 83-86.

[6] Schneider T., Brennan R.L., *A Multichannel Compression Strategy for a Digital Hearing Aid*, Proc. ICASSP-97, Munich, Germany, 1997,pp.411-415.

[7] Menendez-Pidal X., Chen R., Wu D., Tanaka M., *Front-end improvement to reduce stationary & variable channel and noise distortions in continuous speech recognition tasks,* Proc. EUROSPEECH"99, 1999, Vol.6, pp. 2849-2851.

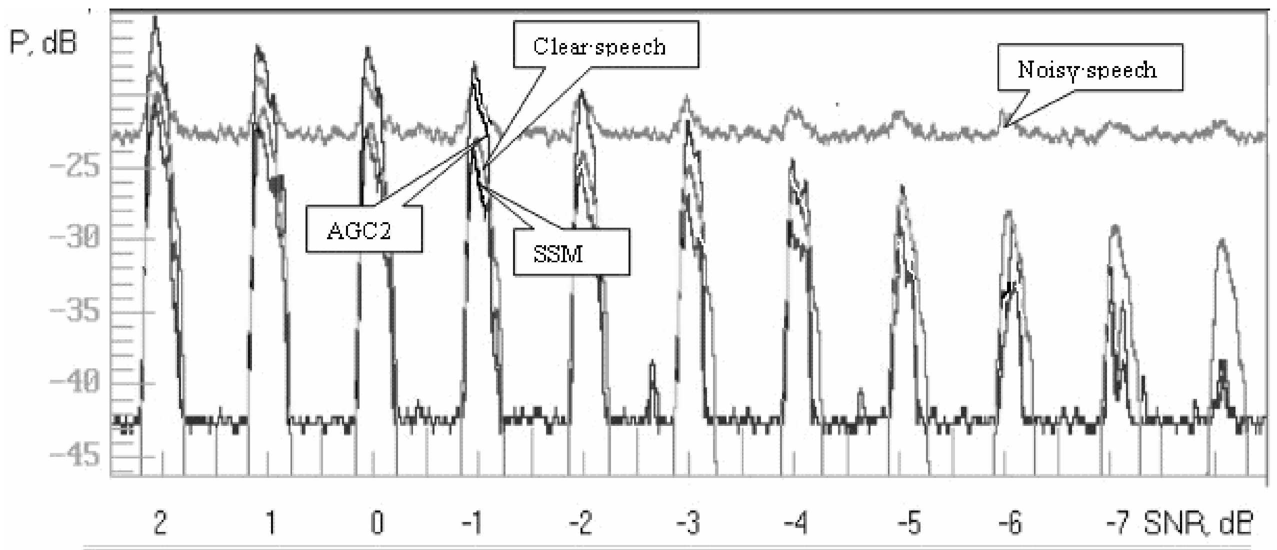

Figure 3. The power of input noisy speech signal (upper curve), clear speech (the third from the top curve), SSM output (bottom curve) (*Go*=-18 dB, *b*=1) and AGC2 output (the second from the top curve) (*Go*=-18 dB, *C*=0.5, *b*=1, *G_max*= 6 dB). The demonstration test signal is the mixture of the white noise and speech signal of the same word "One" with the power decreasing by step 1 dB. The horizontal axe shows the real SNR from +2dB to -8 dB.

Table 2. Average output processed speech signal powers and its real gains for SSM and AGC2.

Noise power P_N = -23 dB.

Clear speech signal power P_S dB|	-9	-11	-13	-15	-17	-19	-21	-23	-25	-27	-29
SNR dB	14	12	10	8	6	4	2	0	-2	-4	-6
INR	6	5	4.2	3.5	3	2.6	2.25	2	1.8	1.6	1.5
P_{SSM} dB	-9	-11	-13	-15	-17	-19	-22	-24	-27	-30	-35
G_{SSM}dB	0	0	0	0	0	0	-1	-1	-2	-3	-6
P_{AGC2} dB	-3	-5	-7	-9	-12	-14	-17	-19	-22	-27	-33
G_{AGC2} dB	6	6	6	6	5	5	4	4	3	0	-4