NOISE SUBTRACTION WITH PARAMETRIC RECURSIVE GAIN CURVES

Klaus Linhard and Tim Haulick*
e-mail: klaus.linhard@daimlerchrysler.com
* e-mail: tim.haulick@vs.dasa.de
DaimlerChrysler Research and Technology, D-89081 Ulm, Germany
* DaimlerChrysler Aerospace, D-89077 Ulm, Germany

ABSTRACT

A well known technique for speech enhancement is spectral subtraction. This technique is attractive due to its simplicity and its low computational cost. We have proposed a simple modification of the filter coefficient calculation in form of a recursion of the previous filter coefficient. Because of this recursion the subtraction rule (gain curve) consists of two different paths, one for increasing signal-to-noise-ratio and another for the decreasing ratio. In this paper we will introduce a parametric version of this recursive approach with two parameters controlling the crossover behaviour of the two paths. Although this recursive version is a very simple algorithm it is easily adjusted to produce no musical noise and additionally to get low speech distortion and/or low residual noise.

10.21437/Eurospeech.1999-575

1. INTRODUCTION

Speech enhancement based on spectral subtraction (- subtracting a noise estimate) has the drawback of generating noise with musical character, the so-called musical noise. Reduction of musical noise is very desirable, but should not be managed sacrificing other issues of speech quality. We may formulate the problem in the following way: First it is necessary to guarantee that there is no musical noise and then we should look for the speech quality in terms of low speech distortion and low residual noise.

There are many papers about successful approaches for eliminating the annoying musical noise, e.g. [2], [4]. We propose a new approach based on standard spectral subtraction and a simple recursion in the calculation of the gain curve. This approach neither needs an additional processing delay nor a complicated estimation procedure. The recursive gain is implemented by feedback of the gain value one time instant before.

We will first summarise the notation and our definition of standard spectral subtraction:

We assume that the speech signal s is additively corrupted by some noise n:

$$x(t) = s(t) + n(t) \quad . \tag{1}$$

Spectral subtraction is usually implemented by multiplying the Fourier transformed input segment X with the real filter coefficients H and getting Y:

$$Y_{k,i} = H_{k,i}\, X_{k,i} \quad . \tag{2}$$

X results from time segments of $x(t)$ which are usually half overlapped and multiplied with a Hanning window. k and i denote the discrete segment index and the discrete frequency index, respectively. Succeeding time segments resulting from inverse transformation of Y are half overlapped and added to give a continuous output time signal $y(t)$.

We will refer to the following standard gain curves ,[1]:

Magnitude subtraction:

$$H_{k,i} = P_b\left[\, 1 - \sqrt{\frac{a}{INR_{k,i}}}\, \right] \;, \tag{3}$$

Power subtraction:

$$H_{k,i} = P_b\left[\, \sqrt{1 - \frac{a}{INR_{k,i}}}\, \right] \;, \tag{4}$$

Wiener approximation:

$$H_{k,i} = P_b\left[\, 1 - \frac{a}{INR_{k,i}}\, \right] \;, \tag{5}$$

with the input-to-noise ratio

$$INR_{k,i} = \frac{|X_{k,i}|^2}{|N_{k,i}|^2} \;, \tag{6}$$

and the noise estimate obtained in speech pauses

$$|N_{k,i}|^2 = (1-\alpha)\,|N_{k-1,i}|^2 + \alpha\,|X_{k,i}|^2 \quad . \tag{7}$$

$P_b[x]$ is the projection: $P_b[x] = x$ if $x \geq b$ and $P_b[x] = b$ else.

These three standard gain curves differ in their attenuation which is a function of the noise estimate, respectively INR. All these curves have a lower bound b on the gain H in common. b is called

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

spectral floor and serves to mask musical noise. An overestimation factor a (≥ 1) is used to account for fluctuations of the spectral noise power and therefore overestimating the noise reduces the musical residual noise. Fig.(1) shows a comparison of the three gain curves. Our approach is not restricted to these curves, other more complicated curves are also possible. For easy tractability we will discuss only the Wiener approximation.

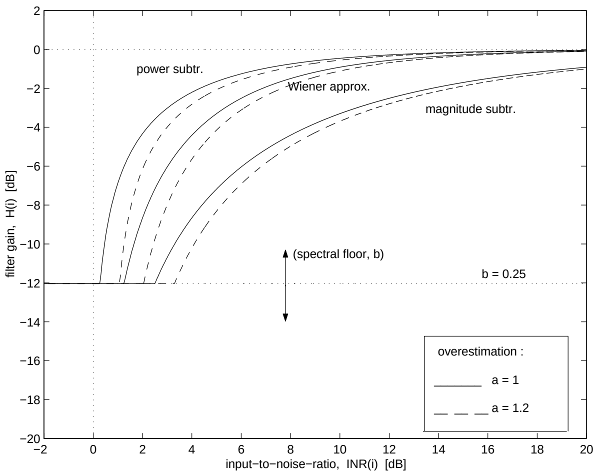

Fig. 1. Standard gain curves of spectral subtraction

2. RECURSIVE GAIN CURVES

Recursion of the gain curves is accomplished with the substitution:

$$INR_{k,i} := INR_{k,i}\ H_{k-1,i} \quad . \tag{8}$$

Eq.(8) may be used to modify the Wiener suppression rule yielding, [5]:

$$H_{k,i} = P_b \left[\, 1 - \frac{a}{INR_{k,i}\ H_{k-1,i}} \,\right] , \tag{9}$$

and with the additional parameter c

$$H_{k,i} = P_b \left[\, 1 - \frac{a}{INR_{k,i}\ (\ (1-c) + c\ (H_{k-1,i} - b))} \,\right] . \tag{10}$$

We will call Eq.(10) the parametric recursive gain curve. Fig.(2) shows the standard Wiener curve Eq.(5) and the two recursive curves Eq.(9) and Eq.(10) with the parameters: $a = 1$; $b = 0.25$; $c = 0.8$. The figures are calculated with slowly increasing *INR* and decreasing *INR* again afterwards (step size is 0.01% of the range of *INR*). In Fig.(2) and all following figures the standard Wiener curve is also plotted for comparsion.

Due to the recursion the gain depends on its previous value which results in a hysteresis, thus

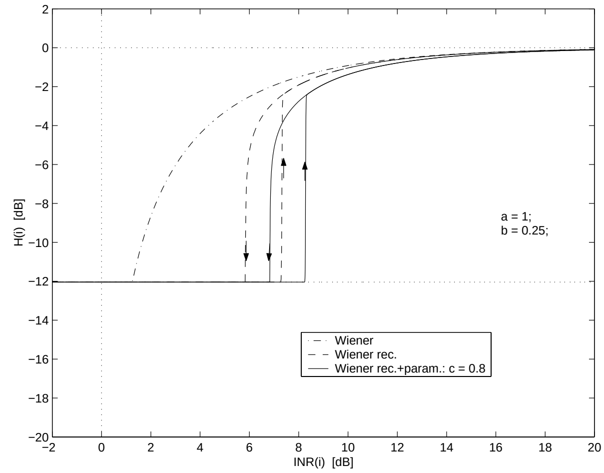

Fig. 2. Recursive gain curves

there exists a crossover area with two different paths (arrows indicate the direction in relation to *INR*).

There are three differences between the standard Wiener curve and the recursive version:

- $H = b$ for an extended range of small values for *INR*, thus in speech pauses small *INR* values are processed with a constant gain value b and no musical noise could be generated.
- Switching over from $H = b$ to $H <\approx 1$ happens suddenly, thus the gain curve shows two almost constant parts, $H = b$ and $H <\approx 1$
- There is a gap between the two switching paths of increasing and decreasing *INR*.

The gap may be interpreted in the way that a high gain value stays a little bit longer if the speech energy just is decreasing. We postulate that this is an advantage in terms of exploiting psychoacoustic postmasking. The speech signal plays the role of the masker, masking the noise (residual noise, musical noise). If speech is present it masks the noise. If speech energy decreases it will need some time (in the order of several 10ms, see e.g. [6]) until noise could be heard again. Thus it would be an unnecessary action to adjust to a new gain value until speech energy is really low.

We may compare our recursion with the decision directed approach of Ephraim and Malah, [2], [3]. For this purpose we simplified the original approach by applying the Wiener suppression rule (see also [5]). The gain values H are calculated with the aid of the so-called a priori and a posteriori signal-to-noise ratios:

$$R_{post,k,i} = \frac{|\ X_{k,i}\ |^2}{|\ N_{k,i}\ |^2} - 1 \ , \tag{11}$$

$$R_{prio,k,i} = (1-\alpha)\ P[R_{post,k,i}] + \alpha\ \frac{|\ H_{k-1,i}\ X_{k-1,i}\ |^2}{|\ N_{k,i}|^2} \ , \tag{12}$$

$$H_{k,i} = \frac{R_{prio,k,i}}{R_{prio,k,i}+1} \quad , \tag{13}$$

where the projection $P[x] = x$ if $x \geq 0$ and $P[x] = 0$ else. α is a weighting parameter. Fig.(3) shows the resulting gain curves for three different values of α. The gap between the forward and backward path depends on α. The comparison of the recursion in Eq.(10) and Eq.(12) shows that the recursion of Eq.(10) depends on H_{k-1} while the recursion in Eq.(12) depends on H_{k-1} and also on X_{k-1}. Further our recursion doesn't need the averaging of Eq.(12). A detailed comparison is beyond the scope of this paper but nevertheless Fig.(3) shows typical characteristics which we also have in our approach.

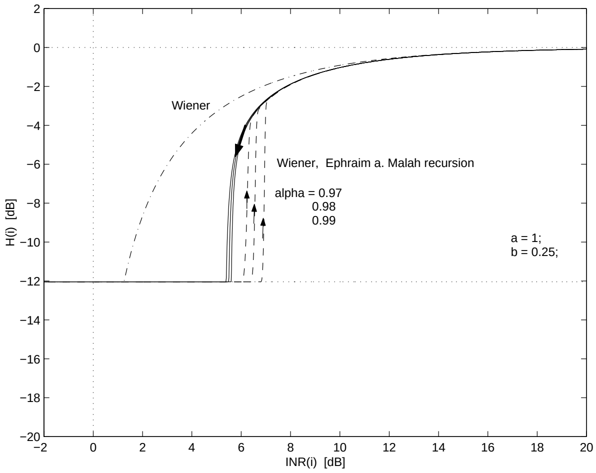

Fig. 3. Gain curve resulting from modified 'Ephraim a. Malah'

Eq.(10) is the parametric recursive suppression rule. It depends on three parameters: a, b and c. Where a is the overestimation factor, b is the spectral floor, and c is a new parameter controlling the gap between the two paths. In the following we will use parameter a to shift the crossover area. Figs.(4) and (5) show a reasonable range for shifting the curve with $0.5 < a < 1.5$ and for changing the gap with $0.75 < c < 0.9$, respectively.

Figs.(6) and (7) show gain curves for the parameter sets:

A) :
$a = 0.5/0.46/0.42/0.38$; $c = 0.75/0.8/0.85/0.9$
B) :
$a = 1.0/0.92/0.84/0.76$; $c = 0.75/0.8/0.85/0.9$

with $b = 0.25$. The gap parameters c are the same in each set, but c is not really independent from the shifting parameter a, thus we adjusted parameter a to get the same falling edge for the different gaps. This way we got four different shifting parameters a

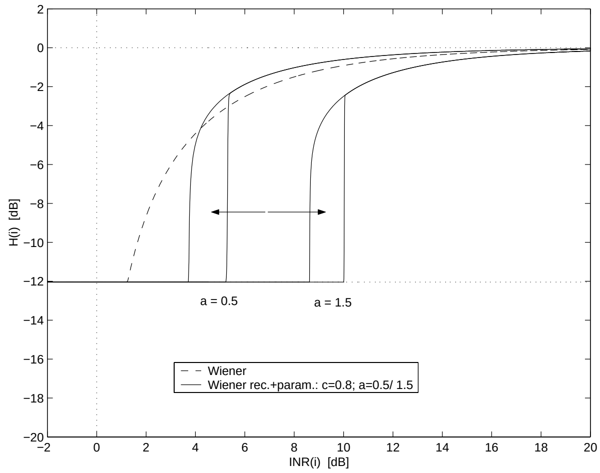

Fig. 4. Shifting the recursive gain curves

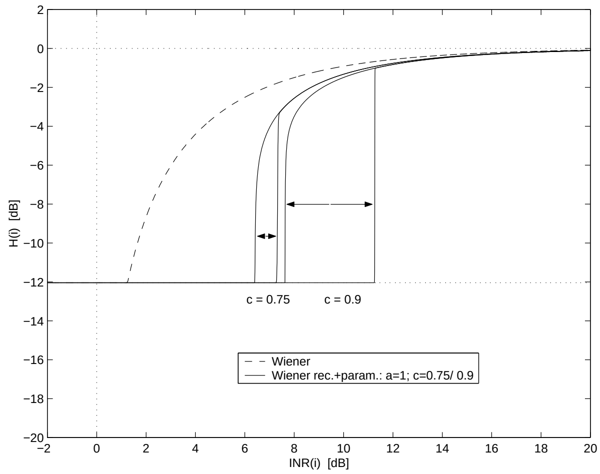

Fig. 5. Changing the gap between the increasing and decreasing path

for each set instead of only one.

3. RESULTS

We performed an informal subjective listening test. The database consists of $\approx 15s$ speech from a male and from a female speaker, each. Road noise from a Mercedes car was added to the speech for the driving situations: $100km/h$ with and without sliding roof partially open. Signal-to-noise ratio is in the range of 0 to $10dB$.

First we used the parameters of set A) and B) and an additional set C) with the same parameters like B) but with spectral floor $b = 0.15$. The spectral floor $b = 0.15$ allows a maximum noise reduction of $20 * log10(0.15) = 16.5dB$, and with $b = 0.25$ the maximum is $20 * log10(0.25) = 12dB$.

After listening to the results with these sets we discarded the examples where musical noise could be heard. Thus from A) only the last subset was

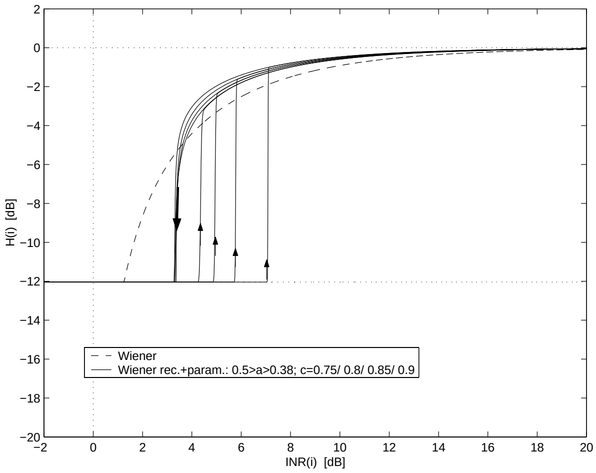

Fig. 6. Different gaps with the same decreasing path at INR=3dB

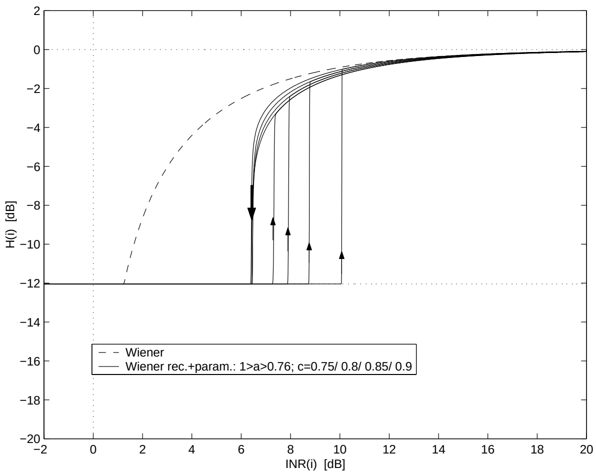

Fig. 7. Different gaps with the same decreasing path at INR=6dB

left, from B) the last three and from C) only the last two ones were left. Now we added several parameter combinations which we assumed would give better results than before but again without musical noise and presented this examples to several persons.

Because of the restricted time available for testing it was not possible to find optimum parameters but there was a clear preference to choose a large gap and a crossover area shifted to the left (e.g.: $a = 0.3$; $c = 0.95$). In terms of overall speech quality $b = 0.25$ was preferred against $b = 0.15$.

4. CONCLUSIONS

We presented a parametric recursive spectral subtraction technique. The recursion may be easily implemented within different subtraction rules. The recursion results in a crossover area for the gain curve. Two parameters allow to control the gap of the crossover area and its position. Informal listening tests indicate that the two parameters are well suited to adjust the algorithm to get good speech quality without musical noise.

References

[1] Vary. P.: "On the Enhancement of Noisy Speech", Proc. of EUSIPCO '83, pp.327-330, 1983

[2] Ephraim, Y.; Malah, D.: "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Trans. on Acoustics, Speech, a. Signal Processing, Vol.ASSP-32, No.6, pp.1109-1121, Dec. 1984

[3] Cappé, O.: "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor", IEEE Trans. on Speech a. Audio Processing, Vol.2, No.2, pp.345-349, April 1994

[4] Linhard, K.; Klemm, H.: "Noise Reduction with Spectral Subtraction and Median Filtering for Suppression of Musical Tones", Proc. of ESCA-NATO Workshop on Robust Speech Recognition for Unknown Communication Channels, pp.159-162,1997

[5] Linhard, K.; Haulick, T.: "Spectral Noise Subtraction With Recursive Gain Curves", Proc. of ICSLP 98, Conf. on Spoken Language Processing, pp.1479-1482,1998

[6] Zwicker, E.; Fastl, H.: "Psychoacoustics - Facts and Models", 2nd Edition, Springer, 1999