SPECTRAL NOISE SUBTRACTION WITH RECURSIVE GAIN CURVES

Klaus Linhard and Tim Haulick

DAIMLER BENZ AG, Research and Technology, Wilhelm-Runge-Str. 11,
D-89081 Ulm, Germany
e-mail: linhard@dbag.ulm.daimlerbenz.com

ABSTRACT

A well known technique for speech enhancement is spectral subtraction. This technique is attractive due to its simplicity and its low computational cost. However, spectral subtraction has the drawback of generating residual noise with a musical character, the so-called musical noise. In this paper, we propose a simple modification of the filter coefficient calculation in form of a recursion of the previous filter coefficient. As indicated by objective and subjective tests, the new recursive enhancement scheme substantially reduces musical noise without a noticeable impairment of the speech signal.

1. INTRODUCTION

A widely used method for speech enhancement is the method of spectral subtraction [1], [2]. Spectral subtraction gained its attraction because of its easy implementation, but it also has some disadvantages, namely speech distortion and residual musical tones. Several researchers reported about improvements in reducing or eliminating the annoying musical tones. One especially successful approach for this task is reported in [4], [5], [6]. This approach is much more mathematically founded than standard spectral subtraction and needs a complicated mechanism for estimating the gain curve. Another approach [7] uses median filtering of the spectra in time direction to remove musical tones. This method is especially simple, but it needs one or two future samples from future short time spectra. The necessary processing delay may be a disadvantage at least for some real time applications. The proposed new approach neither needs processing delay nor a complicated estimation procedure. The potential for eliminating musical tones is gained by a simple recursion of the previous gain curve.

2. SPECTRAL SUBTRACTION AND GAIN CURVES

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, in its simplest form, is 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 $y(t)$.

There are several suggestions of gain curves. In [4], for example, modified Bessel functions are used. For easy tractability we will refer to standard gain curves, (see for example [3]):

Magnitude subtraction:

$$H_{k,i} = \begin{cases} 1 - \sqrt{a \;/\; INR_{k,i}} & ; \quad 1 - \sqrt{a \;/\; INR_{k,i}} > b \\ b & ; \quad \text{(else)} \; , \end{cases} \tag{3}$$

Power subtraction:

$$H_{k,i} = \begin{cases} \sqrt{1 - a \;/\; INR_{k,i}} & ; \quad 1 - a \;/\; INR_{k,i} > b^2 \\ b & ; \quad \text{(else)} \; , \end{cases} \tag{4}$$

Wiener approximation:

$$H_{k,i} = \begin{cases} 1 - a \;/\; INR_{k,i} & ; \quad 1 - a \;/\; INR_{k,i} > b \\ b & ; \quad \text{(else)} \; , \end{cases} \tag{5}$$

with

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

INR is the input-to-noise ratio and the noise estimate $|N_{k,i}|^2$ is obtained in speech pauses:

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

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 $(0.1 < b < .3)$ is also called 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 but a gain curve should have a lower bound b to be suitable for our following recursive gain curves.

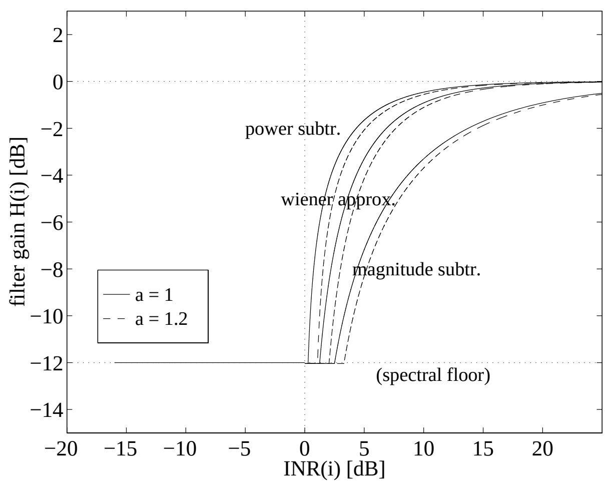

Figure 1: Standard gain curves of spectral subtraction ($b = 0.25$)

3. 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 standard curves and yielding for example for the Wiener suppression rule, Eq.(5):

$$H_{k,i} = P_b\ [\ 1 - \frac{a}{(INR_{k,i}\ H_{k-1,i})}\ ] \quad , \tag{9}$$

where the projection $P_b[x] = x$ if $x \geq b$ and $P_b[x] = b$ otherwise. One result of the recursion is that the gain is fixed to the value b if *INR* is below a threshold value. This switch over to b happens suddenly, thus the gain curve practically consists of two different parts, a part with constant value b and a part with a gain value depending on the choice of one of the standard curves (or some other gain curve).

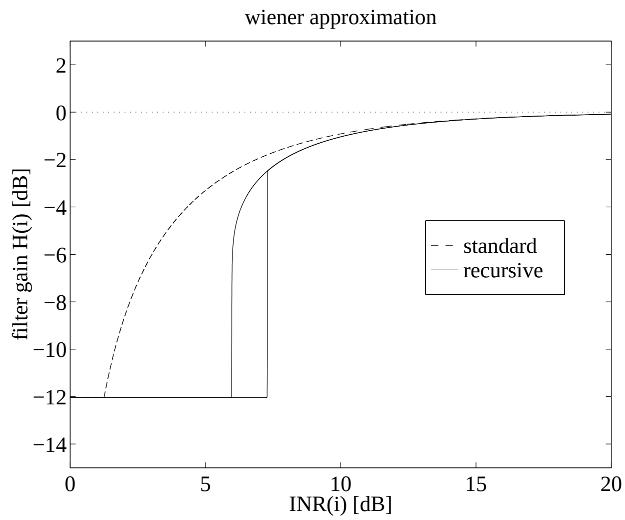

Figure 2: Recursive gain curve for Wiener approx. (increasing and decreasing *INR*, $b = 0.25$)

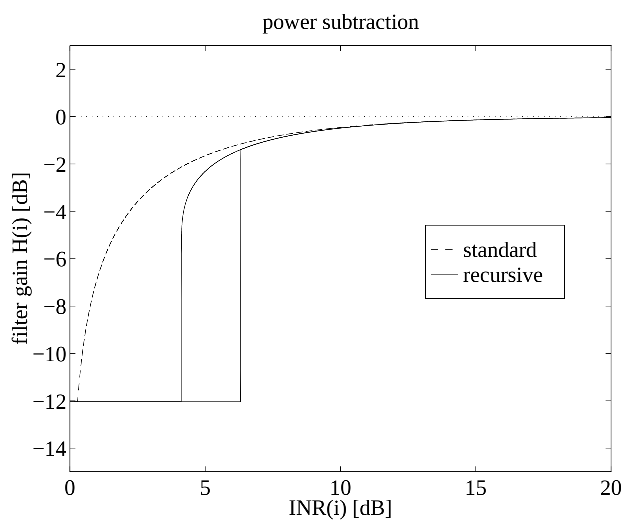

Figure 3: Recursive gain curve for power subtraction (increasing and decreasing *INR*, $b = 0.25$)

Figs. (2) and (3) show the resulting gain curves in case of the Wiener approximation or power subtraction, respectively. (The figures are calculated with slowly increasing *INR* and decreasing *INR* again afterwards.)

Due to the recursion, the gain depends on its previous value which results in a hysteresis. Fig.(4) shows a comparison of the standard power subtraction rule and its recursive version. Fig.(4) shows the spectral input and estimated speech output magnitude, $|X|$ and $|Y|$, for an input signal with the first 20 frames noise and the next 20 frames with noise plus a constant speech component 10 dB above the average noise level. While the output $|Y|$ with standard power subtraction is following or even emphasizing short time spectral peaks of $|X|$ in the noise only situation, the recursive version is reproducing the input at a lower level (naturally sounding background noise). The recursive version is able to follow the onset speech component with only very short delay.

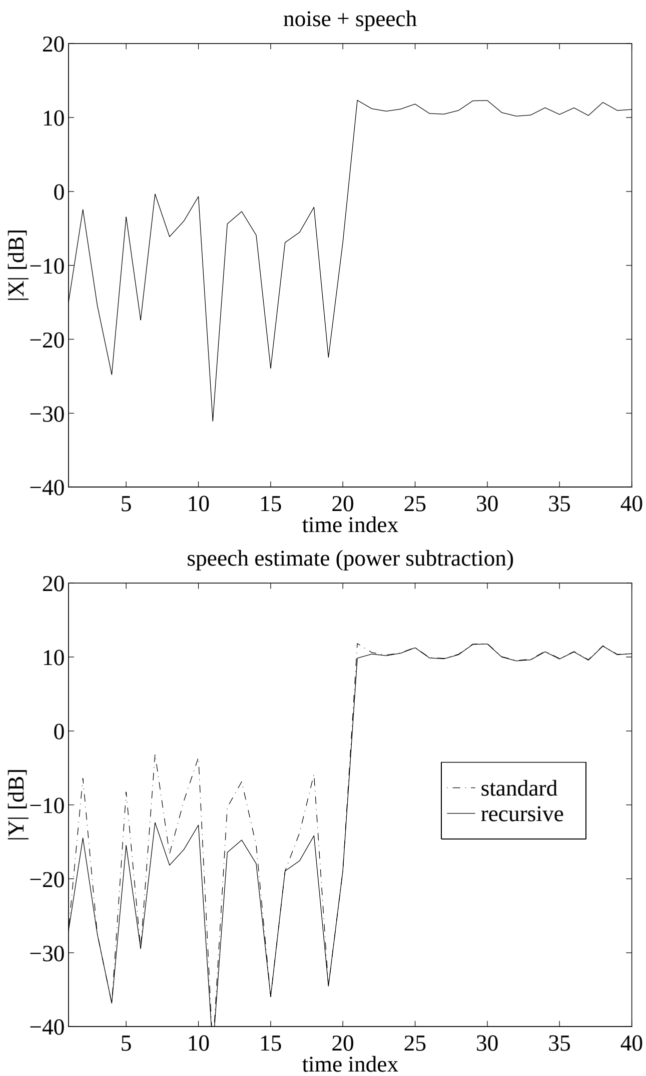

Figure 4: Noisy input $|X|$ and estimated speech output $|Y|$ applying standard and recursive power subtraction rules ($b = 0.25$)

In comparison to our simple approach the recursive gain estimation of [4], [5] and [6] is accomplished with the aid of the so-called a priori and a posteriori signal-to-noise ratios

$$R_{post,k,i} = \frac{|X_{k,i}|^2}{|N_{k,i}|^2} - 1 \ , \tag{10}$$

$$R_{prio,k,i} = (1-\alpha)\ P[R_{post,k,i}] + \alpha\ \frac{|H_{k-1,i}\ X_{k-1,i}|^2}{|N_{k,i}|^2} \ , \tag{11}$$

where the projection $P[x] = x$ if $x \geq 0$ and $P[x] = 0$ otherwise. The weighting parameter α is generally chosen very close to 1 (for ex. 0.98). Following the simplifications in [6], the Wiener approximation gain is calculated, (Eq.(13) in [6]):

$$H_{k,i} = \frac{R_{prio,k,i}}{R_{prio,k,i} + 1} \ , \tag{12}$$

in contrast to the standard Wiener curve (Eq.(5) with $a = 1$):

$$H_{k,i} = P_b\ [\frac{R_{post,k,i}}{R_{post,k,i} + 1}] \ . \tag{13}$$

Eq.(12) is the Wiener approximation with the 'older' complicated form of the recursion and Eq.(13) is just the standard Wiener approximation without recursion.

Comparison of Eqs.(12) and (13) shows that the performance improvements are based on the replacement of $R_{post,k,i}$ with the smoothed estimate $R_{prio,k,i}$. Within speech pauses, $R_{prio,k,i}$ is highly smoothed and in case of high level speech input $R_{prio,k,i}$ and therefore the gain curve is determined by the recursion product $H_{k-1,i}\ X_{k-1,i}$. This recursion product introduces a short delay and therefore is the source of a small reverberation.

Applying our notation to Eq.(13) yields:

$$H_{k,i} = P_b\ [\frac{INR_{k,i} - 1}{INR_{k,i}}] \ . \tag{14}$$

Applying our recursion, the substitution Eq.(8), to Eq.(14) yields:

$$H_{k,i} = P_b\ [\ 1 - \frac{1}{INR_{k,i}\ H_{k-1,i}}\ ] \ , \tag{15}$$

$$H_{k,i} = P_b\ [\ 1 - \frac{|N_{k,i}|^2}{|X_{k,i}|^2 H_{k-1,i}}\ ] \ . \tag{16}$$

The recursion product $X_{k,i}\ H_{k-1,i}$ is de-coupled in time (notice the different time indicies k and $k-1$) and therefore the source of reverberation is less in comparison to the original approach of [4].

4. RESULTS

We performed several calculations and informal subjective listening tests comparing our simple recursive approach to the more complicated one, Eqs.(10) and (11), the median filtering method [7] and also a psychoacoustic based approach [8]. The new approach compares well with the others in terms of high noise reduction, low speech distortion and improved overall speech quality. All the above methods reduce musical noise to a low level or eliminate it. In terms of noise reduction, the recursive approaches are capable to outperform the others at the price of more speech distortion. Up to now the median filtering method seemed to have the lowest computational complexity. The new recursive scheme is still easier for implementation yet yielding a comparable or sometimes better performance than the other approaches.

5. REFERENCES

[1] Boll, S.F.: "A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech", Proc. on ICASSP, pp.200-203, 1979

[2] Boll, S.F.: "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Vol. ASSP-27, No. 2:113-120, April 1979

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

[4] 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

[5] 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

[6] Azirani, A.A.; le Bouquin Jeannès, R.; Faucon, G.: "Speech Enhancement Using a Wiener Filtering Under Signal Presence Uncertainty", pp. 971-974

[7] 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

[8] Haulick, T.; Linhard, K.; Schrögmeier, P.: "Residual Noise Suppression Using Psychoacoustic Criteria", EUROSPEECH 97, 22-25 Sept. 1997, pp. 1395-1398