COMPARISON OF ONE– AND TWO–CHANNEL NOISE–ESTIMATION TECHNIQUES

Joerg Meyer, Klaus Uwe Simmer and Karl Dirk Kammeyer

University of Bremen, FB–1, Dept. of Telecommunications
P.O. Box 330 440, D–28334 Bremen, Germany, email: meyer@comm.uni-bremen.de

ABSTRACT

In this paper, we compare several one–channel and two–channel noise–estimation techniques. We focus on their estimation features in a non-stationary noisy environment while a speech signal is present, as this is one of the unsolved problems for spectral subtraction algorithms. All one–channel solutions make use of the different statistics of speech and unwanted noise. The two–channel algorithms use the spatial characteristics of the noise field in order to estimate the power spectral densities (PSD). First, we will briefly describe several existing algorithms, then we will introduce a new one which is related to the one proposed by Gierl [1].

1. INTRODUCTION

In this contribution, we describe several techniques to estimate noise while a speech signal is present. Estimating noise is essential for spectral subtraction algorithms. Using a voice activity detector (VAD) does not work well, as the VAD assumes that the noise is stationary during speech intervals. Furthermore, there are no robust one-channel VADs up to now. To overcome this problem, several one–channel noise–estimation techniques without VAD have been introduced during the last few years [2, 3, 4, 5, 6]. Another way of handling this problem is to use a two-channel approach to estimate the noise [1, 7]. In the first part of our work we will give a short summary of the existing techniques. Secondly, a new algorithm is introduced and we show that this algorithm theoretically can estimate the noise perfectly, if the spatial noise characteristic is known. Finally, we will compare all algorithms in terms of their abilities to estimate the noise of a mixed signal, consisting of slowly amplitude-modulated noise added to clean speech.

2. ALGORITHMS

2.1. One–Channel Algorithms

- Voice Activity Detector (VAD)
- Direct Estimation (DE) [3]
- Modified Direct Estimation (MDE) [3]
- Threshold Direct Estimation (TDE) [5]
- Histogram Technique (HT) [5]
- Minimum Statistics, Martin (MSM) [2]
- Minimum Statistics, Doblinger (MSD) [4]
- Iterative Wiener Filter (IWF) [6]

All one–channel noise estimation techniques use recursive formulas in subbands (FFT-based) to estimate the noise. This recursion can be interpreted as a Welsh periodogram. The algorithms differ from each other in the rules to update these formulas. The most complex algorithms are the MSM and the HT techniques. These algorithms additionally use the different probabilities of noise and speech to separate the two.

2.2. Two–Channel Algorithms

- Channel Subtraction (SUB) [1]
- Cross–Correlation Based (CC) [7]
- Modified Channel Subtraction (NEW)

The two–channel approaches use the different spatial correlations of noise and speech to estimate the noise. They assume that the speech is highly correlated at different sensors and that the noise source is either diffuse or spatially separated from the speech source.

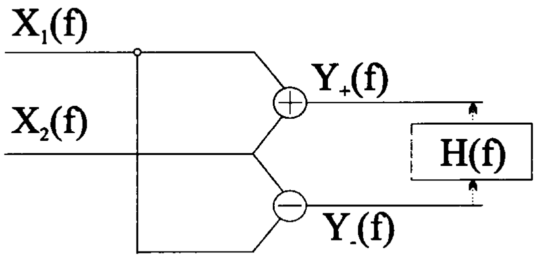

Figure 1: Blockdiagram

2.3. Modified Subtractive Algorithm

If we assume that the speech signal is perfectly correlated and there is no time-delay between the signals on the microphones, there are different ways of estimating the noise. One solution is to block the speech by subtracting signals of different sensors (see figure 1). The residual signal $Y_-(f)$ does not contain any speech, but noise only. This noise has to be transformed by a filter $H(f)$ in order to compute the noise estimation of the added signals $Y_+(f)$. Gierl proposed a system-identification algorithm to estimate the filter. It can be shown that this filter only depends on the real part of the complex coherence-function of the noise

$$\Gamma_{X_1X_2}(f) = \frac{P_{X_1X_2}(f)}{\sqrt{P_{X_1X_1}(f)P_{X_2X_2}(f)}} \quad , \qquad (1)$$

where $P(f)$ denotes power spectral densities (PSD).

To compute the transformation filter $H(f)$ we first compute the PSD of $Y_+(f)$

$$P_{Y_+Y_+} = X_1X_1^* + X_2X_2^* + X_2X_1^* + X_1X_2^* \quad . \quad (2)$$

If we assume that no speech signal is present and that the PSD of the noise is $P_{NN} = P_{X_1X_1} = P_{X_2X_2}$ it follows using equation (1):

$$\begin{aligned} P_{Y_+Y_+} &= 2P_{NN} + 2\Re\{X_2X_1^*\} \qquad (3) \\ &= 2P_{NN}(1 + \Re\{\Gamma_{X_1X_2}\}) \end{aligned}$$

The PSD for Y_- is similar:

$$\begin{aligned} P_{Y_-Y_-} &= X_1X_1^* + X_2X_2^* - X_2X_1^* - X_1X_2^* \\ &= 2P_{NN} - 2\Re\{X_2X_1^*\} \\ &= 2P_{NN}(1 - \Re\{\Gamma_{X_1X_2}\}) \qquad (4) \end{aligned}$$

Thus the filter is

$$H(f) = \frac{1 + \Re\{\Gamma_{x_1x_2}(f)\}}{1 - \Re\{\Gamma_{x_1x_2}(f)\}} \quad . \qquad (5)$$

In order to estimate the coherence of the noise we still need a coarsly working VAD. However, we are no longer dependent on time stationarity but only on the spatial stationarity of the noise. It can be shown that spatial stationarity is more likely to be given (for example in cars) [1, 8].

3. SIMULATION

To compare the algorithms under different noise situations we first simulated three rooms with varying reverberation time (τ_{60} = 0ms, 150ms and 1000ms) by using the image method described by Allen and Berkley[9]. In order to get mixed signals we convolved one speech sentence and white Gaussian noise with their respective room impulse response and mixed them at different time-varying levels. In the first test we have simulated decreasing noise. We started at 0dB signal to noise ratio (SNR) and increased it with velocities from 0dB/s (stationary noise) to 5dB/s. In a second test increasing noise is simulated. The beginning SNR is 10dB and decreases with the same velocities (see figure 2). The criterion for the noise estimation performance is the absolute error of the estimation which is normalized to the absolute error of the VAD algorithm. It is computed as

$$\frac{\sum_{f=0}^{fs}\sum_{k=wb}^{we}\left|N - \hat{N}_{algo}(k,f)\right|}{\sum_{f=0}^{fs}\sum_{k=wb}^{we}\left|N - \hat{N}_{VAD}(k,f)\right|} \quad , \qquad (6)$$

where wb and we denote the beginning and the end of the sentence. The true power spectral density (PSD) of the noise is denoted by N, and $\hat{N}_{algo}$ is the estimated noise of the selected algorithm. This error criterion has the same weighting for under- and overestimation of the noise, but the effects for noise reduction algorithms are different: there is either residual noise or signal cancellation.

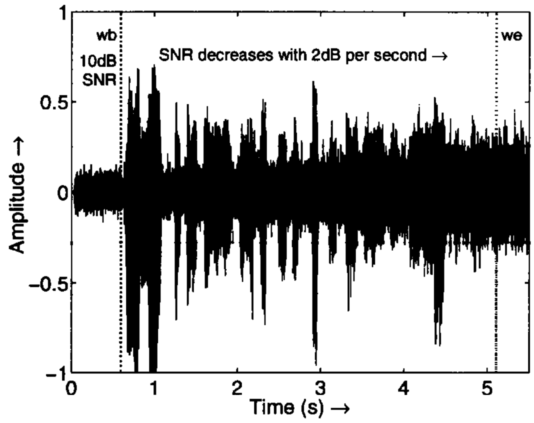

Figure 2: Mixed input signal for the increasing noise experiment

4. RESULTS

In the figures 3 to 6 the results for the simulation with the reverberation time of τ_{60} = 150ms are shown.

Due to the normalization the results for the VAD are always one, but the absolute error for the VAD increases dramatically. It is 3 times larger in the decreasing noise test at 5db/s and 115 times larger in the increasing case. This shows, that no one–channel technique gives good results if noise is increasing. The two–channel algorithms outperform the one–channel solution, but the noise estimation performance is poor, too, if the noise increases fast. The results for τ_{60} = 0ms and τ_{60} = 1000ms are comparable. Only the CC method gives better results for larger reverberation times. These results confirm that the one–channel algorithms are independent of the noise condition.

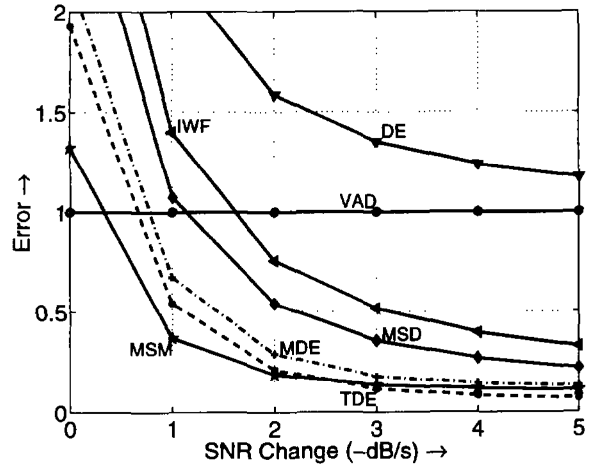

Figure 3: One–channel algorithms, decreasing noise, normalized to the VAD Error

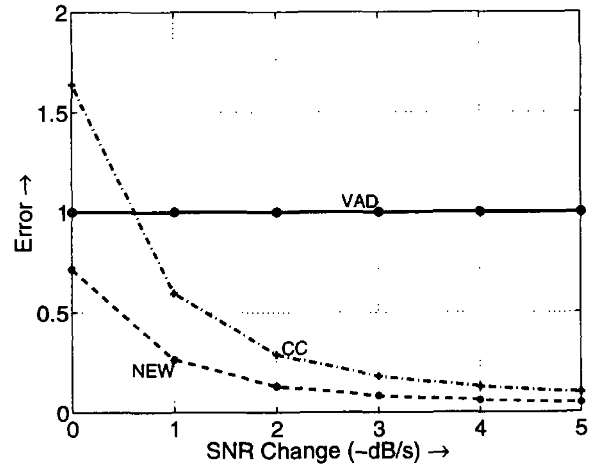

Figure 4: Two–channel algorithms, decreasing noise, normalized to the VAD Error

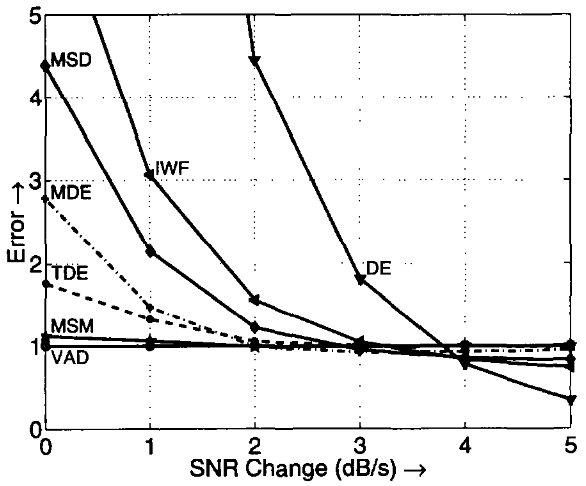

Figure 5: One–channel algorithms, increasing noise, normalized to the VAD Error

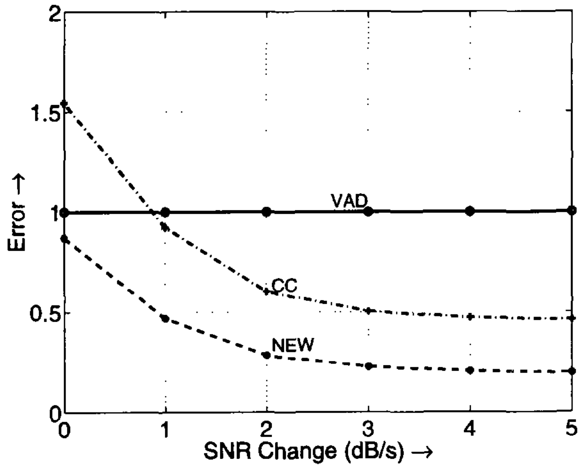

Figure 6: Two–channel algorithms, increasing noise, normalized to the VAD Error

4.1. Comments on the algorithms

DE: The DE algorithm fails in all cases because of the strong signal cancellation. This is a typical problem for this type of algorithm.

MDE: Due to the stronger restriction of the attack time in the recursive formulas of this algorithm, the signal cancellation effect is less than that of the DE. This algorithm is the best one–channel choice, if the complexity of the algorithm has to be as low as possible.

TDE: The TDE algorithm shows a better noise estimation performance than the MDE, but this algorithm is not as robust as the MDE.

HT: Due to its complexity, the HT algorithm (not shown in the figures) is not to be recommended. The noise estimation performance is comparable to that of the TDE. Only in the stationary case this algorithm performs noticeably better.

MSM: The MSM is the best one–channel algorithm for noise estimation. It is robust and has very good noise estimation features for decreasing noise. Its only disadvantage is its complexity and its need for much memory.

MSD: The MSD algorithm has a tendency to cancel the signal. The parameter choice is more difficult.

IWF: The IWF has a different structure and has to be seen in the complete noise reduction scheme. For noise estimation this algorithm is unsuitable, but the noise reduction and the residual speech quality is comparable to that of other one–channel algorithms.

SUB: The SUB (not shown in the figures) has the same features and results as our new approach, because of the close relation of the two algorithms (see NEW).

CC: The CC algorithm works best in an uncorrelated noise field, but this noise condition cannot be assumed for speech acquisition. There is always a high coherence for low frequencies, dependent on the microphone distances, and furthermore there are often correlated frequencies, due to resonance frequencies of noise sources.

NEW: Our NEW algorithm (and SUB) gives best results for all cases. It works with all SNRs and in the case of increasing noise. The only drawback is, there is a need for a coarsly working VAD. But for two input channels quite good VADs are presented by [10].

5. CONCLUSION

In this contribution we have given a comparison of known noise estimation techniques. Furthermore, a theoretically motivated new approach was introduced. The results show that the one–channel estimation techniques do not work well when noise increases. Only the two-channel techniques give accurate results in all cases.

6. REFERENCES

[1] S. Gierl, *Geräuschreduktion bei Sprachübertragung mit Hilfe von Mikrofonarraysystemen.* PhD thesis, Universität Karlsruhe, 1990.

[2] R. Martin, "Spectral Subtraction Based on Minimum Statistics," in *European Signal Processing Conference (EUSIPCO-94)*, (Edinburgh, UK), pp. 1182–1185, September 1994.

[3] L. Arslan, A. McCree, and V. Viswanathan, "New Methods for Adaptive Noise Suppression," in *Proc. IEEE Int. Conference Acoustic, Speech and Signal Processing, ICASSP–95*, (Detroit, Michigan), pp. 812–815, Mai 1995.

[4] G. Doblinger, "Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands," in *European Conference on Speech Technology and Communication, EUROSPEECH 95*, (Madrid, Spain), pp. 1513–1516, September 1995.

[5] H. G. Hirsch and C. Ehrlicher, "Noise Estimation Techniques for Robust Speech Recognition," in *Proc. IEEE Int. Conference Acoustic, Speech and Signal Processing, ICASSP-95*, (Detroit, Michigan), pp. 153–156, Mai 1995.

[6] P. Sovka, P. Pollak, and J. Kybic, "Extended Spectral Subtraction," in *European Signal Processing Conference (EUSIPCO-96)*, (Trieste, Italy), September 1996.

[7] M. Dörbecker and S. Ernst, "Combination of Two–Channel Spectral Subtraction and Adaptive Wiener Post–Filtering for Noise Reduction and Dereverberation," in *European Signal Processing Conference (EUSIPCO-96)*, (Trieste, Italy), September 1996.

[8] J. Meyer and K. U. Simmer, "Multi–Channel Speech Enhancement in a Car Environment Using Wiener Filtering and Spectral Subtraction," in *Proc. IEEE Int. Conference Acoustic, Speech and Signal Processing, ICASSP–97*, (Munich, Germany), April 1997.

[9] J. B. Allen and Berkley D. A., "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[10] R. Le Bouquin, "Enhancement of Noisy Speech Signals: Application to Mobile Radio Communication," *Speech Communication*, vol. 18, pp. 3–19, 1996.