A Family of Coherence-Based Multi-Microphone Speech Enhancement Systems

Quang Hung PHAM, Pavel SOVKA

Dept. of Circuit Theory, Czech Technical University in Prague, Technická 2, 166 27 Praha 6, Czech Republic

hung@amber.feld.cvut.cz, sovka@feld.cvut.cz

Abstract. *This contribution addresses the problem of additive noise reduction in speech picked up by a microphone in a noisy environment. Two systems belonging to the family of coherence-based noise cancellers are presented. Suggested systems have the modular structure using 2 or 4 microphones and suppress non-stationary noises in the range of 4 to 17 dB depending on the chosen structure and noise characteristics. The common properties are acceptable noise suppression, low speech distortion and residual noise.*

Keywords

Noise reduction, speech enhancement, coherence function, non-stationary noises, spectral subtraction.

1. Introduction

The problem of noise reduction or speech enhancement can be found in various applications of speech processing. Especially, speech recognizers working in adverse conditions or speech pre-processing for hearing impaired often include subsystems insuring noise robustness.

Several techniques have already been developed to solve the problems of noise reduction and speech enhancement. These techniques may be divided into one- and multi-microphone systems. Each technique is applicable for certain types of noise environments such as reverberations in an office, noises in a car cab, and so on. But almost all techniques use two basic principles for noise attenuation: a subtraction of an estimated noise floor or filtering an input signal.

This work deals with noise reduction techniques using the coherence function that have been proposed in the last few years. These techniques are analyzed and experimentally compared with the aim of further modification and combination to increase their robustness. As a result, modified techniques are proposed as pre-processing systems for cellular phones in a running car.

2. Problem Definition

Let us take speech $s[n]$ contaminated by an additive noise $n[n]$ to get noisy speech

$$x[n] = s[n] + n[n]. \tag{1}$$

or in the spectral domain

$$X(f) = S(f) + N(f). \tag{2}$$

When speech and noise are uncorrelated then spectral density of noisy speech $x[n]$ has the form

$$P_{XX}(f) = Pss(f) + P_{NN}(f), \tag{3}$$

where $P_{SS}(f)$, and $P_{NN}(f)$ are spectral densities (PSD) of speech and noise, respectively.

The aim is to get the estimation of speech (enhanced speech) $\hat{s}[n]$ or its spectrum $\hat{S}(f)$. This problem is known as noise reduction or speech enhancement. Now, two basic principles used in speech enhancement methods will be given.

2.1 Two Basic Principles of Noise Reduction

As mentioned before, there are many techniques for noise reduction. These techniques use two basic principles for noise attenuation.

A. Spectral subtraction

The very simple and common principle is spectral subtraction and its modifications. The main idea of spectral subtraction is to subtract the estimation of a background noise spectral density $\hat{P}_{NN}(f)$ from the instantaneous input spectrum according to

$$\tilde{X}_a(f) = |X(f)| - \hat{P}_{NN}(f). \tag{4}$$

Discussion:

1. The phase needed for the enhanced speech reconstruction $\hat{s}[n]$ is taken from the input signal spectrum $X(f)$. Then

$$\widetilde{X}(f) = \widetilde{X}(f)_a \exp(j \arg(X(f))) \quad (5)$$

2. Spectral density $\hat{P}_{NN}(f)$ is estimated in pauses when speech is absent. Thus a reliable voice activity detector (VAD) is required. The difference between an instantaneous noise spectrum *N*(*f*). and its long-term estimation $\hat{P}_{NN}(f)$ introduces errors known as residual noises. The more non-stationary background noise the more the stronger residual noises appear.

3. The PSD $\widetilde{X}_a(f)$ must be non-negative thus two types of realization of eq. (4) are used

- half-way rectification (HWR)

$$\widetilde{X}_a(f) = \max(|X(f)| - \hat{P}_{NN}(f), 0), \quad (6)$$

- half-way rectification (HWR)

$$\widetilde{X}_a(f) = |(|X(f)| - \hat{P}_{NN}(f))|. \quad (7)$$

Both procedures cause speech distortion. Therefore the first proposed spectral subtraction method [1] has been modified by many authors as in [1], and [2] to reduce this distortion. Despite these efforts, this technique has its own sizeable residual noise that is called a musical or residual noise. Modifications of spectral subtraction try to reduce the speech distortion and residual noises using some sort of temporal or ensemble averaging of an input signal or enhanced speech. As shown in [3] the addition of noise to speech causes increasing the mean and variance of spectrum *X*(*f*). The spectral subtraction performed according to equations (4) and (5) recovers the original mean of clean speech leaving the variance unchanged. That means the background noise spectrum is suppressed but residual noises can be heard in the enhanced speech. If some sort of averaging is applied to enhanced speech, e.g. [4], [5], [6] residual noises can be attenuated at the expense of further speech distortion. Therefore one aim of this contribution is to use a proper procedure attenuating residual noises and at the same time decreasing the speech distortion caused by applying eq. (6) or (7), and a temporal or ensemble averaging of enhanced speech. We suggest also not to apply the frequently used noise floor masking residual noises [6], [7] because this approach decreases the final noise reduction.

It can be concluded that the spectral subtraction principle requires a reliable noise estimation procedure and VAD. Therefore a lot of effort has been spent on the development of these two parts.

B. Wiener or matched filtration

Another principle used is Wiener filtration or more generally matched filtration. These approaches decompose an input signal into non-overlapping frequency bands and attenuate the bands containing strong noises according to

$$\widetilde{X}(f_i) = X(f_i) G(f_i), \quad (8)$$

where f_i is the center frequency of the *i*-th frequency band and $G(f_i)$ are known as spectral gains forming usually a real transfer function using the Wiener or matched filter.

Comparison of both principles:

- Wiener or matched filtration does not require PSD rectification (eq. (6) or (7)) but residual noises and speech distortion are still present and comparable with the distortion caused by spectral subtraction.
- Spectral subtraction is a highly non-linear process due to rectification. Filtration is a purely linear process and therefore some additional objective measures evaluating speech distortion and noise attenuation can be used. These measures give additional information about system performance (see section 5).

2.2 Multi-Microphone Systems for Noise Reduction

The principles described are used in both one- or multi-microphone systems. Multi-microphone schemes use these principles in various modifications and represent robust techniques for noise reduction. In this case the spatial information can be used to control a directivity pattern and thus to increase the efficiency of the whole system. Studies for fixed directivity pattern or the adaptive directivity pattern can be found e.g. in [8], [9], or [11]. More inputs enable use of the coherence function between two input signals $x_1[n]$ and $x_2[n]$ defined by [12]

$$C_{X_1X_2}(f) = \frac{P_{X_1X_2}(f)}{\sqrt{P_{X_1X_1}(f) P_{X_2X_2}(f)}}, \quad (9)$$

where $P_{X_1X_2}(f)$ is the cross power spectral density of input signals $x_1[n]$ and $x_2[n]$. The magnitude of the coherence function is equal to one for fully correlated signals and zero for uncorrelated signals. A speech signal is more correlated than noises in a car cabin thus the coherence function can be used as the sensitive measure for the discrimination between speech and noise or between coherent and diffusive noises. The use of this function does not require any compensation of a time delay between $x_1[n]$ and $x_2[n]$.

Having the described advantages of the coherence function in our mind we have focused on systems using these functions, e.g. [13], [14], [15]. These systems are very robust and they are able to suppress diffusive as well as coherent noises effectively with a small number of microphones. In order to optimize these systems it is advantageous to use a modular structure in which all parts of the system are well separated. The suggested solution using two or four microphones will be described in the following sections.

3. Suggested Solution of Using Two-Microphone Systems

To insure a high level of modularity it is advantageous to use a general structure of multi-microphone systems depicted in Fig. 1:

- The first part contains at least 2 microphones followed by A/D converters, an orthogonal transformation converting the signal to the frequency domain, a voice activity detector (VAD), and a time delay compensation block. This part ensures the proper directivity pattern of the system but it is not solved in this contribution.
- The second part performs background noise estimation $\hat{P}_{NN}(f)$. In the case of the two-microphone method, the best results can be obtained by using coherence-based estimators. This part is analyzed in this contribution.
- The third part cancels the noise from the input signal. This part can be implemented using various methods. This paper describes only three methods: spectral subtraction as the simplest one, and the adaptive Wiener filter updated by Ephraim-Malah [18] and Akbari-Azirani [19] approaches as the standard methods. While the Ephraim-Malah approach leads to a very efficient and robust method, the Akbari-Azirani algorithm represents a simpler but computationally more efficient version of Ephraim-Malah method.

The fourth part is responsible for the attenuation of residual noises arising from errors in the background noise estimation.

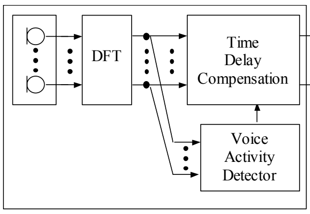

Fig. 1. General structure of a multi-microphone noise reduction system

Originating from the structure due to M. Dörbecker [16] we focus on the simplified structure shown in Fig. 2. This structure represents the second part of the noise reduction system from Fig. 1 for the case when two microphones are used. The system has four independent parts. Thus the almost arbitrary combination of algorithms for the blocks *Noise Estimation*, *Noise Reduction*, and *Post-filtering* can be analyzed and the used algorithms can be optimized almost independently. At the same time this structure can be easily generalized to a structure with four-microphones (see next section).

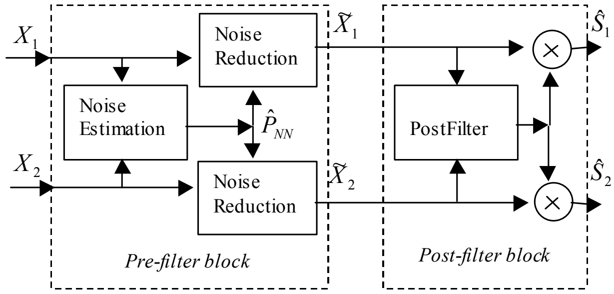

Fig. 2 Simplified two-microphone noise reduction system

3.1 Noise Estimation Methods

As pointed out above the estimation of background noise is the key of all noise reduction systems. This contribution is focused on coherence-based estimators requiring at least two microphones and ensuring a high level of robustness. In real environments there are typically two types of background noise. Schematically, a diffusive noise is typical for large rooms with reverberations while a coherence noise is present in small rooms or a car cabin. A typical coherent noise is the engine noise in a car interior. Unfortunately, there is no coherence-based algorithm enabling the estimate of both types of noise simultaneously. One possible solution is to use a filter for noise reduction switching between the Wiener filter (for diffusive noises) and a coherence filter (for coherence noise) [15]. The solution used here is to switch between two algorithms for the noise spectrum estimation while using one Wiener filter for noise reduction. Both algorithms for the noise spectrum estimation will be described in the following section.

3.1.1 Estimation of the Diffusive Noise Spectrum

One possible approach is the Dörbecker coherence-based estimator [16]

$$\hat{P}_{NN}(f) = \sqrt{\hat{P}_{X_1X_1}(f)\,\hat{P}_{X_2X_2}(f)} - \left|\hat{P}_{X_1X_2}(f)\right|, \tag{10}$$

where $\hat{P}_{X_1X_1}(f)$, $\hat{P}_{X_2X_2}(f)$ and $\hat{P}_{X_1X_2}(f)$ are power spectral densities and cross power spectral density, respectively.

The estimation of spectral densities $\hat{P}_{XiXj}(f)$ for k^{th} segment is based on a first-order IIR filter

$$\hat{P}_{X_iX_j}(f,k) = \beta\hat{P}_{X_iX_j}(f,k-1) + (1-\beta)X^*_i(f)X_j(f), \tag{11}$$

where β is a forgetting factor and $i,j = 1,2$..

Algorithm conditions summary: These estimates are valid if signals $s_1[n]$ and $s_2[n]$ are correlated while noises $n_1[n]$ and $n_2[n]$ are mutually uncorrelated and uncorrelated with signals $s_i[n]$.

3.1.2 Estimation of the Coherent Noise Spectrum

Another noise estimation method was proposed by Simmer [12]. This method estimates a background noise using an addition and subtraction of two input signals defined as follows: a signal $Y_+(f)$ is the addition of two input signals $X_1(f)$ and $X_2(f)$ while a residual signal $Y_-(f)$ as the subtraction of these two input signals.

The PSD of a noise contained in the signal $Y_+(f)$ may be expressed by the transformation of the PSD of the residual signal [17]

$$\hat{P}_{NN}(f) = H(f)\, P_{Y_-Y_-}(f) \,. \tag{12}$$

The transfer function $H(f)$ is given by

$$H(f) = \frac{1 + \Re\{C_{X_1X_2}(f)\}}{1 - \Re\{C_{X_1X_2}(f)\}} \,, \tag{13}$$

where $\Re\{C_{X_1X_2}(f)\}$ denotes the real part of the coherence function $C_{X_1X_2}(f)$ between both input signals.

The estimation of the short-time coherence function is obtained by the PSDs and cross-PSDs of the input signals

$$\hat{C}_{X_1X_2}(f,k) = \frac{\hat{P}_{X_1X_2}(f,k)}{\sqrt{\hat{P}_{X_1X_1}(f,k)\hat{P}_{X_2X_2}(f,k)}} \,, \tag{14}$$

where k is a segment index. The PSDs $\hat{P}_{X_1X_1}(.)$, $\hat{P}_{X_2X_2}(.)$, and $\hat{P}_{X_1X_2}(.)$ can be estimated using equation (11).

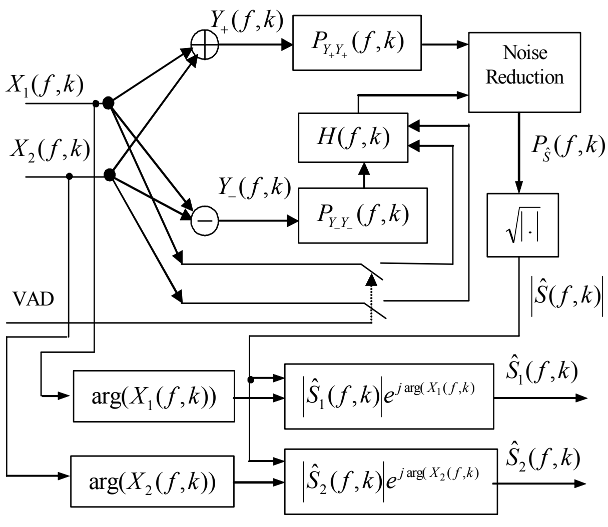

Fig. 3. Two-microphone coherence-noise reduction system creating *Pre-filter Block* from Fig. **2**

Algorithm conditions assumption: These estimations are valid under the assumptions that the input speech signals are correlated, the *PSD*s of the noises are equal, and the estimation $\hat{P}_{NN}(f)$ is performed in pauses when no speech is present in the input signal. Thus this technique needs a VAD. On the other hand, it doesn't require the time stationarity of the noise but only the spatial stationarity of the noises.

The choice of VAD depends on the application. One-microphone VADs can be used for higher signal-to noise ratio (SNR). These VADs are simple and give relatively precise detection of speech activity intervals. When the SNR is low (below 0 dB) it is necessary to use two-microphone VADs often using the coherence function. If a coherence-based VAD and a coherence–based noise estimation procedure are used then it is possible to save computational costs. The coherence-based VADs are more robust than one-microphone VADs although the localization of speech activity segments is not as precise for coherence-based VADs than for one-microphone VADs [20], [21].

The whole structure for the reduction of coherent noises is depicted in Fig. 3. The block *Noise Reduction* is realized using later described noise reduction algorithms.

3.2 Used Noise Reduction Algorithms

The heart of noise reduction systems is the noise reduction algorithm. Many various sorts of these algorithms have been published. This contribution, as stated before, is focused on spectral subtraction as the simplest method, the Ephraim-Malah algorithm as the standard one and its efficient implementation known as the Akbari-Azirani algorithm. A short description of the chosen algorithms will be given now.

3.2.1 Noise Attenuation by Spectral Subtraction

Spectral subtraction was analyzed in detail in the preceding text. Here it can be summarized that it is a method for the restoration of the power or the magnitude spectrum of a signal observed in additive noise, through the subtraction of an estimate of the average noise spectrum from the noisy signal spectrum.

Spectral subtraction is very simple therefore it may be implemented even for systems without powerful hardware. The cost for its simplicity, as mentioned above, is significant speech distortion and residual noise caused by rectification, random noise variations and VAD errors.

Spectral subtraction using magnitude spectrum and full-wave rectification was used in our system.

3.2.2 Noise Attenuation by Filtration

The spectral gain $G(f)$ used in equation (8) is computed by the Ephraim and Malah method [18] called the MMSE and by Akbari-Azirani method [19].

The MMSE method based on modeling speech and noise spectra as statistically independent Gaussian random variables is one of the best one-microphone noise suppression methods. The spectral gain depends on two parameters: the a priori and a posteriori signal to noise ratio. This approach has less musical noise than spectral subtraction while significantly suppressing color noises.

The Akbari-Azirani method is based on Wiener filtering under uncertainty of signal presence. It can be seen as a computationally efficient modification of the MMSE method with a comparable noise suppression effect. This approach allows easier implementation of noise reduction systems using digital signal processors.

3.3 Post-Filter Reduction of Residual Noises

The fourth part (see Fig. **2**) of suppressing residual noise is performed by the Wiener filter [16]

$$W(f,k)=\frac{\left|P_{\tilde{X}_1\tilde{X}_2}(f,k)\right|^2}{\left(P_{\tilde{X}_1\tilde{X}_1}(f,k)+P_{\tilde{X}_1\tilde{X}_2}(f,k)\right)^2}, \tag{15}$$

where $\tilde{X}_1(f,k)$ and $\tilde{X}_2(f,k)$ are short-time spectra generated by the *Noise Reduction* block.

4. Generalization to Four-Microphone Systems

The described two-microphone noise reduction system can be improved by extending it to four input microphones. Such a system uses the spatial information in a more efficient way and therefore has better performance. The way to use this information depends on the geometry of the microphone array and on the combination of the basic building blocks of the system. One possible approach is to create all possible pairs of input signals and to filter them through *Pre-filter Blocks* from Fig. 2. Output signals from these blocks can be then averaged. This operation suppresses residual noise, and at the same time, enlarges the level of noise reduction. When input signals are synchronized then the speech distortion caused by this averaging is negligible. This approach is efficient when noise in all microphones is uncorrelated.

The suggested modification increases the required number of operations by about one half of the number of all possible signal pairs. But the computational costs can be decreased by the proper selection of the signal pairs which contribute to the final noise reduction is the most important. One factor for choosing which pairs should be left out is the geometry of a used microphone array. Analyses and simulations showed that in the case of

- non-symmetric geometry: the microphone closest to a speaker serves as a reference microphone and it must be used in all signal combinations
- symmetrical microphone array: it is possible to use arbitrary one half of all possible combinations.

In these both cases the noise reduction is better than for the two-microphone method while the number of needed operations increases slightly (for details see [11]).

5. Experiments and Results

5.1 Signal Database

Experiments were performed on the database of real speech and noise signals picked up by four microphones in various types of quiet or running cars. Two types of a microphone geometry were used - the symmetric geometry with equidistant spaced microphones in line or in a square and non-symmetric one (details see [11]). In order to evaluate the objective criteria for noise reduction and speech distortion clean signals were added to noises. Four types of noises were used – stationary and non-stationary in spectrum, and stationary and non-stationary in energy. Gain K used for the scaling of noise before adding it to speech was constant or changing with time.

5.2 Criteria

Outputs of the noise reduction systems were evaluated by using following criteria:

A. Signal-to-noise ratio enhancement SNRE

$$SNRE=10\log\frac{\sum_{k=1}^{N}s^2[k]}{\sum_{k=1}^{N}\left(s[k]-\hat{s}[k]\right)^2}-10\log\frac{\sum_{k=1}^{N}s^2[k]}{\sum_{k=1}^{N}n^2[k]}, \tag{16}$$

where N is number of samples in a segment, $s[k]$ and $n[k]$ are speech and noise, $x[k]$ is the speech signal disturbed by an additive noise, $\hat{s}[k]$ is an output speech signal enhanced by a noise reduction system. The second term on the right hand of equation (16) is the signal-to-noise ratio.

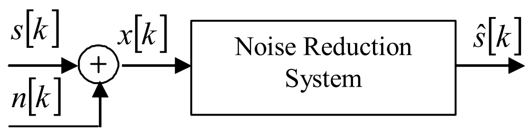

Fig. 4. The diagram for derivation of the SNR and SNRE criteria

SNRE measures the efficiency of a speech enhancement system but its correlation with the subjective evaluation of a listener is low.

B. Segmental *SNRE*

$$SSNRE=\frac{1}{M}\sum_{i=1}^{M}SNRE[i], \tag{17}$$

where M is the number of segments, $SNRE[i]$ is $SNRE$ for the i^{th} segment. $SNRE$ shows better correlation with subjective hearing.

Other used criteria as the speech distortion (*SD*), and the noise reduction (*NR*) and their segmental versions (*SegSD* or *SegNR)* require the noise reduction system to be the linear system. In this case the weights W_F of the adaptive system (master) can be copied into two slave systems with the inputs $s[k]$ and $n[k]$ (see Fig. 5). Then we assess the quality of the speech enhancement system in terms of the distortion of the speech signal and noise reduction by passing the clean speech signal and the noise through the adaptive filters with coefficients W_F (slave systems in Fig. 5). Using enhanced speech $\hat{s}[k]$, distorted speech $\tilde{s}[k]$ and reduced noise $\tilde{n}[k]$ we can estimate the speech distortion and noise reduction according to equations (18) to (21) as follows.

C. Speech distortion *SD*

$$SD = 10\log\frac{\sum_{k=1}^{N}(\tilde{s}[k]-s[k])^2}{\sum_{k=1}^{N}s^2[k]} \tag{18}$$

D. Noise reduction *NR*

$$NR = 10\log\frac{\sum_{k=1}^{N}n^2[k]}{\sum_{k=1}^{N}\tilde{n}^2[k]}, \tag{19}$$

where $\tilde{s}[k]$ and $\tilde{n}[k]$ are clean speech signal or noise, respectively, filtered by the adaptive slave filters with coefficients W_F.

E. Segmental *SegSD* or *SegNR*

$$SegSD = \frac{1}{M}\sum_{i=1}^{M}SD[i], \tag{20}$$

$$SegNR = \frac{1}{M}\sum_{i=1}^{M}NR[i], \tag{21}$$

where M is the number of segments, $SD[i]$, $NR[i]$ are the SD, NR of the i^{th} segment, respectively.

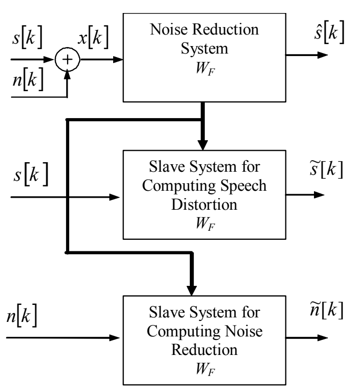

Fig. 5. The structure for the generation of the signals required by the *NR* and *SD* criteria

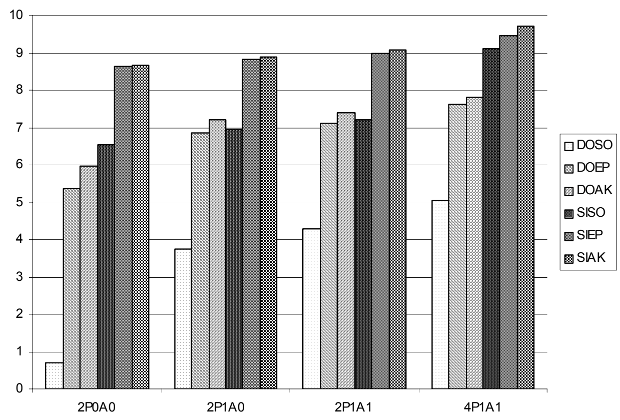

Fig. 6. *SSNRE* of selected structures for stationary noise Abbreviations of the structures are created by the following rules. "DO" and "SI" stand for the Dörbecker and Simmer noise estimation methods. "SO", "EP", and "AK" stand for Spectral Subtraction, MMSE, and Akbari noise reduction methods. The numbers 2 and 4 equal the number of microphones. "P1" and "P0" determine whether a system includes *Post-filter* block or not. "A1" and "A0" determine whether outputs are averaged or not. For example, the first column (DOSO, 2P0A0) shows the *SSNRE* value of the two-microphone Dörbecker structure with Spectral Subtraction, without *Post-filter* block, and without averaging.

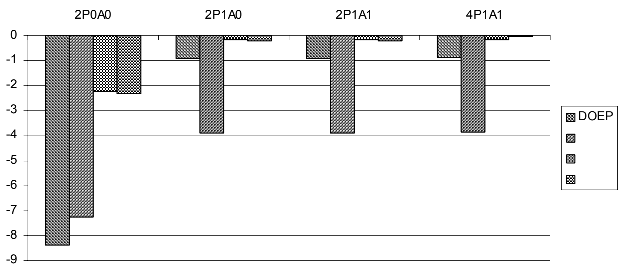

Fig. 7. Speech Distortion of speech enhancement structures Abbreviations of the structures are described in Fig. **6**.

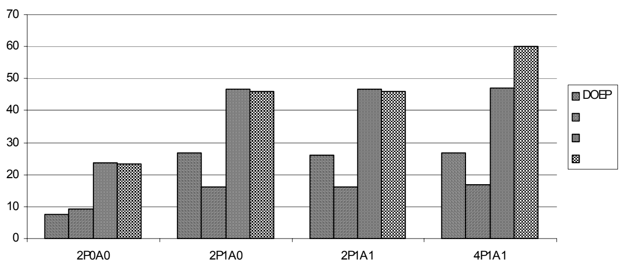

Fig. 8. Noise Reduction of speech enhancement structures Abbreviations of the structures are described in Fig. **6**.

5.3 Results

Typical results for stationary noise are shown in Fig. 6, 7, and 8. Results for non-stationary noise were similar. SegSD and SegNR cannot be evaluated for structures with spectral subtraction since it is not linear. By comparing the obtained results the following conclusions can be made:

- Structures with the Simmer noise estimation method (sec. 3.1.2) have better SSNRE values than the structure with the Dörbecker noise estimation method (sec. 3.1.1).
- The spectral subtraction method distorts spectral characteristics of the speech signal, thus the averaging of output signals has an important influence. Four-microphone structures with averaging have much better SSNRE values than two-microphone structures. A four-microphone with the Dörbecker noise estimation method gains about 4.3 dB against the two-microphone structure. A four-microphone structure with the Simmer noise estimation method has a roughly 3.3 dB better SNRE value than a two-microphone structure.
- Structures with the Ephraim-Malah method and Akbari-Arizani noise reduction methods (sec. 3.2.2) have comparable SSNRE values and prove much better than the spectral subtraction method. Four-microphone structures with these noise reduction methods behave 1-2 dB better than two-microphone structures. Two-microphone structures with the Simmer noise estimation method and the Aphraim-Malah or the Akbari-Arizani methods give a better SSNRE value than the four-microphone structure with Simmer noise estimation method and spectral subtraction.
- The last rows in Fig. 7 and 8 show that the four-microphone structure with the Simmer noise estimation method using the Akbari-Arizani noise reduction method, with a post-filter and averaging of the output signal has the lowest speech distortion value and the highest noise reduction value. So this structure should be also the best structure by the SSNRE criterion and is confirmed by the results in Fig. 6. Of course, this is the result of this particular example and the specifically chosen weighting factor in implementing the Ephraim and the Akbari noise reduction methods. Our informal listening tests showed that the output speech signal is perceptually acceptable and noise is almost suppressed.

6. Conclusion

Robust two- and four-microphone coherence-based methods were suggested and verified.

The typical signal-to-noise ratio enhancement noise varies from 4 to 9 dB depending on the type of structure and noise reduction algorithm used. Chosen structure and algorithms ensure low speech distortion and residual noises.

Computational and memory requirements of the described systems are low enough to enable their implementation on one signal processor TMS320C30 and to ensure real-time processing.

Acknowledgement

This work was supported by the grant GA 102/02/0124 Voice Technologies for Support of Information Society and the project MSM 212300014 Research in the Area of Information Technologies and Communications.

References

[1] BOLL, S.F. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 1979, vol.ASSP-27, no.2, p. 113-120.

[1] POLLÁK, P., SOVKA, P. Spectral subtraction and speech/pause detection. Internal research report #R95-1, CTU-Faculty of Electrical Engineering, Prague, 1995.

[2] DAVÍDEK, V., SOVKA, P., ŠIKA, J. Real-time implementation of spectral subtraction algorithm for suppression of acoustic noise in speech. In *Proceedings of the 4th European Conference on Speech Communication and Technology,* EUROSPEECH '95, Madrid, pp. 141-144, September 1995.

[3] VASEGHI, S.V: *Advaced signal processing and digital noise reduction.* Wiley Teubner, New York, 1997.

[4] KANG, G.S., FRANSEN L.J. Quality improvement of LPC-processed noisy speech by using spectral subtraction. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 1977, vol. ASSP-37, no.6, pp. 939-942.

[5] LIMM,J.S.: *Speech enhancement*. Prentice-Hall, New Jersey, 1983.

[6] BEROUTI, M., SCHWARTZ, R., MAKHOUL, J. Enhancement of speech corrupted by acoustic noise. In *Proc. of IEEE Conf. on Acoustics, Speech and Signal Processing*, 1979, pp. 208-211.

[7] CAPPÉ, O. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE,* 1994, pp.345-349.

[8] DÖRBECKER, M.: Speech enhancement using small microphone arrays with optimized directivity. IWAENC 97, 1997, pp. 100-103.

[9] MÁŠA, P. Měření zpoždění mezi signály. Doktorská disertační práce, FEL ČVUT, Praha, 2002.

[10] SIMMER, K. U., KUCZYNSKI, P., WASILJEFF, A. Time delay compensation for adaptive multi-channel speech enhancement systems. In *Proceedings URSI ISSSE,* Paris, 1992.

[11] PHAM, Q. H. Optimization and Implementation of Multi-channel Speech Enhancement Methods. Ph.D. thesis, FEE-CTU in Prague, 2002.

[12] BENDAT, J., S., PIERSOL, A. *Engineering applications of correlation and spectral analysis.* John Wiley & Sons, Inc., New York, 1980.

[13] BOUQUIN, R., L. Enhancement of noisy speech signals. *Application to mobile radio communication. Speech Communication*, 1996, vol. 18, pp. 3-19.

[14] MAHMOUDI, D., DRYGAJLO, A. Combined Wiener and coherence filtering in wavelet domain for microphone array speech enhancement. In *Proc. Of Int. Conf. On Acoustics and Speech Processing* ICAASSP'98, Seatle 1998, pp. 385-388.

[15] RODRIGUEZ, G., J., at al. Coherence-based sub-band decomposition for robust speech and speaker recognition in noisy

and reverberant rooms. In *Proc. Of the ASA Int. Conf. On Acoustics* ICA'01, Roma 2001.

[16] DÖRBECKER, M., ERNST, S. Combination of two-channel spectral subtraction and adaptive Wiener post filtering for noise reduction and deverberation. IWAENC 97, 1997.

[17] MEYER, J., SIMMER, K. U., KAMMEYER, K. D. Comparison of one- and two-channel noise estimation techniques. IWAENC 97, 1997.

[18] EPHRAIM, Y., MALAH, D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 1984, vol. ASSP-32, no. 6, p.1109-1121.

[19] AKBARI, A. A., BOUQUIN, R. L. J., BOUQUIN, G. Speech enhancement using a Wiener filtering under signal presence uncertainty. EUSIPCO 96, Trieste, September 1996.

[20] SOVKA, P., POLLÁK, P. The Study of Speech/Pause Detectors for Speech Enhancements Methods. In *Proceedings of the 4th European Conference on Speech Communication and Technology* - EUROSPEECH'95, Madrid 1995, pp.1575-1578.

[21] MARZINZIK, M., KOLLMEIER, B. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE trans. On Speech and Audio Processing*, 2002, vol.10, no.2, pp. 109-118.