A ROBUST GENERALIZED SIDELOBE CANCELLER EMPLOYING SPEECH LEAKAGE MASKING

Adam Borowicz

Faculty of Computer Science, Bialystok University of Technology, Białystok, Poland

Abstract: A novel speech enhancement method based on generalized sidelobe canceller (GSC) structure is presented. We show that it is possible to reduce audible speech distortions and preserve residual noise level under acoustic model uncertainties. It can be done by constraining a speech leakage power according to masking phenomena and conditional minimizing the residual noise power. We implemented the proposed approach using a simple delay-and-sum beamformer model. Finally a comparative evaluation of the selected methods is performed using objective speech quality measures. The results show that the novel method outperforms conventional one providing lower speech distortions.

Keywords: GSC, psychoacoustics, speech enhancement

1. Introduction

A major objective of the speech enhancement is to reduce environmental noise while preserving speech intelligibility. In a context of the multichannel methods the dereverberation and interference suppression is also expected. The most commonly used dereverbaration methods are beamforming techniques [2]. The key idea of the beamforming is to process the microphone array signals to listen the sounds coming from only one direction. Particularly the noise reduction can be implicitly achieved by avoiding noise directions. The linearly constrained minimum variance (LCMV) algorithm has been originally proposed by Frost [4] and it is probably the most studied beamforming technique since then. It minimizes beamformer output variance subject to the set of linear equations that ensure a constant gain in a specified listening direction. The minimum variance distortion-less (MVDR) method [11] can be considered as a special case of the LCVM approach. Another popular technique is generalized sidelobe canceller [5] [12]. The noisy signal domain is split into two orthogonal subspaces where the dereverberation and noise suppression can be performed separately.

Advances in Computer Science Research, vol. 11, pp. 17-29, 2014.

In order to work reasonably well in the reverberant environments, classical beamforming techniques often require a system model identification i.e. knowledge of the acoustic room impulse responses or its relative ratios. These parameters can be fixed or estimated adaptively, however in general it is a difficult task. In addition the beamforming methods are usually very sensitive to the model uncertainties. Recently, much efforts have been made to reformulate the multichannel speech enhancement problem so that the noise reduction can be achieved without performing speech dereverberation [7]. However these methods are out of scope of this article.

The proposed system is based on the GSC beamformer. We directly assume the presence of the system model uncertainties, which results in the estimation errors (speech leakage effect) and thus increased speech distortions. Instead to minimize these errors we propose to use perceptual properties of the auditory system (simultaneous masking phenomena) to make the speech distortions inaudible. In particular, it is observed that for a given spectral power level, there is a masking threshold so that any interferer below this threshold becomes unnoticed. A similar strategy has been proved to be useful in several single channel methods [3] but according our best knowledge it was not used in a field of the multichannel speech enhancement.

2. Notation

Consider an array of N microphones with arbitrary geometry and single speech source $s(t)$ located inside reverberant enclosure. The observation signal at nth microphone is given by:

$$x_n(t) = a_n(t) * s(t) + v_n(t) = y_n(t) + v_n(t), \tag{1}$$

where $*$ denote a convolution operator, a_n is a room acoustic impulse response from the source speech signal to the nth microphone and $y_n(t)$, $v_n(t)$ are the clean speech and noise components received at nth microphone.

The multichannel systems are often implemented in the frequency-domain using the discrete Fourier transform (DFT). The samples are processed on frame-by-frame basis using analysis window of the length M. Let $X_n(\omega)$, $A_n(\omega)$, $S(\omega)$, $Y_n(\omega)$ and $V_n(\omega)$ denote the DFTs of $x_n(t)$, $a_n(t)$, $s(t)$, $y_n(t)$ and $v_n(t)$ respectively. For sufficiently large M (compared to the length of the room impulse response), we can approximate the model (1) as follows [5]:

$$\mathbf{x}(\omega) = \mathbf{a}(\omega)S(\omega) + \mathbf{v}(\omega) = \mathbf{y}(\omega) + \mathbf{v}(\omega), \tag{2}$$

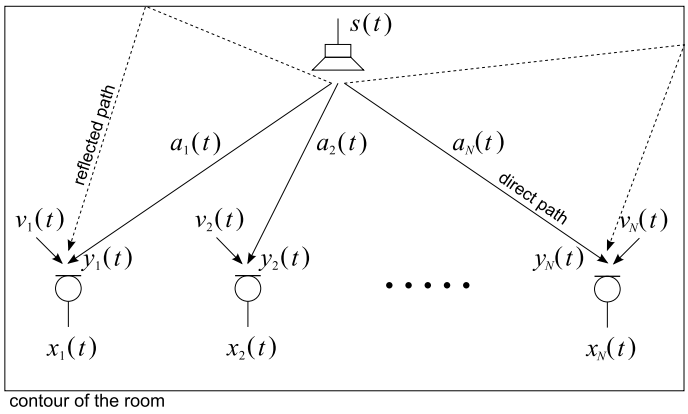

Fig. 1. Graphic illustration of the multimicrophone signal model (1).

where

$$\begin{aligned}\mathbf{x}(\omega) &= [X_1(\omega), X_2(\omega), \ldots, X_N(\omega)]^T,\\ \mathbf{a}(\omega) &= [A_1(\omega), A_2(\omega), \ldots, A_N(\omega)]^T,\\ \mathbf{y}(\omega) &= [Y_1(\omega), Y_2(\omega), \ldots, Y_N(\omega)]^T,\\ \mathbf{v}(\omega) &= [V_1(\omega), V_2(\omega), \ldots, V_N(\omega)]^T.\end{aligned} \tag{3}$$

For example, a correlation matrix for an arbitrary vector $\mathbf{z}(\omega)$ is defined as: $\mathbf{R_{zz}}(\omega) = E\{\mathbf{z}(\omega)\mathbf{z}^H(\omega)\}$, where $E\{.\}$ is an expectation operator and superscript H denotes conjugate transpose. We assumed that the speech and noise processes are wide-sense stationary and uncorrelated, i.e.: $\mathbf{R_{xx}}(\omega) = \mathbf{R_{yy}}(\omega) + \mathbf{R_{vv}}(\omega)$.

3. Speech enhancement: GSC and speech leakage masking

The signal model that we use here (1) can also be presented in the graphical form (Fig. 1). In the case of the frequency domain implementation, our aim is to estimate complex spectrum of the source speech signal, i.e. $S(\omega)$ (then the signal $s(t)$ is obtained from $S(\omega)$ via inverse DFT). The most straightforward way is to apply a linear filter $\mathbf{h}(\omega)$ to observation vector $\mathbf{x}(\omega)$ for each frequency bin:

$$\hat{Y}(\omega) = \mathbf{h}^H(\omega)\mathbf{x}(\omega). \tag{4}$$

Above formula can be viewed as the frequency domain implementation of the finite-impulse-response (FIR) filter. The derivation of the optimal filter $\mathbf{h}(\omega)$ depends on some criteria which we will investigate in the next subsections.

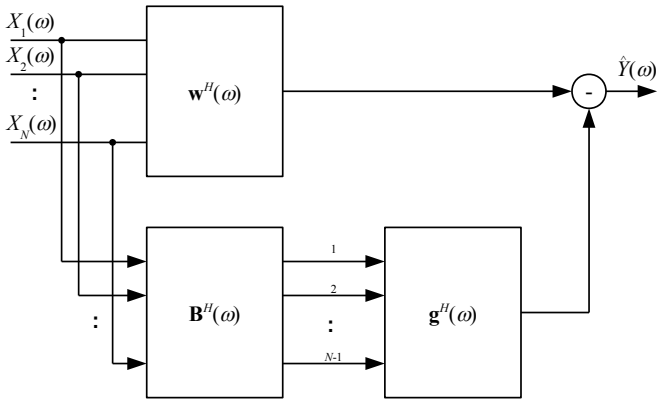

Fig. 2. Block diagram of the GSC beamformer.

3.1 Generalized sidelobe caceller

The GSC approach assumes that the filtering for each channel can be performed in two orthogonal subspaces. It can be expressed mathematically using decomposition of the weighting vector:

$$\mathbf{h}(\omega) = \mathbf{w}(\omega) - \mathbf{B}(\omega)\mathbf{g}(\omega) \tag{5}$$

where $\mathbf{w}(\omega)$ is a steering vector of size N, and $\mathbf{B}(\omega)$ is a blocking matrix of size $N \times (N-1)$ that spans the null space of $\mathbf{A}(\omega)$. The corresponding block diagram of the GSC beamformer is depicted in Fig. 2.

The objective of the GSC approach is to find optimal noise cancellation vector $\mathbf{g}(\omega)$ of size $N-1$. It can be done by solving the following (unconstrained) optimization problem:

$$\min_{\mathbf{g}(\omega)} E\{|\mathbf{w}^H(\omega)\mathbf{v}(\omega) - \mathbf{g}^H(\omega)\mathbf{B}^H(\omega)\mathbf{v}(\omega)|^2\}. \tag{6}$$

Note that this is equivalent to minimizing average residual noise power at the GSC output. An explicit solution for (6) is multichannel Wiener filter [5]:

$$\mathbf{g}_{\mathrm{W}}(\omega) = [\mathbf{B}^H(\omega)\mathbf{R}_{\mathbf{vv}}(\omega)\mathbf{B}(\omega)]^{-1}\mathbf{B}^H(\omega)\mathbf{R}_{\mathbf{v}}(\omega)\mathbf{w}(\omega). \tag{7}$$

Although the GSC and the LCMV beamformers are equivalent, the GSC approach have some interesting interpretation. Note, that the objective of the first vector $\mathbf{w}(\omega)$ is to perform dereverberation on the signal $\mathbf{x}(\omega)$, while the objective of the second

component $\mathbf{B}(\omega)\mathbf{g}(\omega)$ is to suppress the interferences and additive noise. It is worthwhile to note that computationally efficient, adaptive implementations are preferred [5]. However in our experiments we use non-recursive implementation for simplicity.

3.2 Speech leakage constrained method

A major drawback of the GSC beamformer is a high sensitivity to model uncertainties. In fact the performance of the GSC method is affected by the accuracy of the steering vector and blocking matrix estimates. Unfortunately these parameters depend on true channel transfer functions which are usually unknown. Although they can be roughly estimated using second-order statistics [5], [11], in general it is a difficult task. Similarly assuming a simpler acoustic model, can also result in the estimation errors. For example the delay-and-sum beamformer is reliable only in less-reverberant environments.

In our approach, we assume a presence of the estimation errors in the model, explicitly. The output of the GSC beamformer can be decomposed as follows:

$$\hat{Y}(\omega) = \hat{S}(\omega) - \hat{S}_{\mathrm{N}}(\omega) + \hat{V}(\omega) - \hat{V}_{\mathrm{N}}(\omega) \tag{8}$$

where

$$\begin{aligned} \hat{S}(\omega) &= \mathbf{w}^H(\omega)\mathbf{a}(\omega)S(\omega), \\ \hat{V}(\omega) &= \mathbf{w}^H(\omega)\mathbf{v}(\omega), \\ \hat{S}_{\mathrm{N}}(\omega) &= \mathbf{g}^H(\omega)\mathbf{B}^H(\omega)\mathbf{a}(\omega)S(\omega), \\ \hat{V}_{\mathrm{N}}(\omega) &= \mathbf{g}^H(\omega)\mathbf{B}^H(\omega)\mathbf{v}(\omega). \end{aligned} \tag{9}$$

are the beamformer speech component, beamformer noise component, speech leakage and noise reference respectively. If steering vector $\mathbf{w}(\omega)$ is estimated inaccurately, the speech component contains reverberations. Similarly, if $\mathbf{B}^H(\omega)\mathbf{a}(\omega) \neq 0$, the speech signal leakages to the noise cancellation loop i.e. $\hat{S}_{\mathrm{N}}(\omega) \neq 0$, which results in the cancellation of the speech components at the output of the GSC beamformer. It is difficult to improve dereverberation efficiency, however we can minimize the speech leakage effect at expense of some residual noise increase.

Let's define average power of residual noise and speech leakage respectively at the output of the GSC beamformer:

$$\begin{aligned} \varepsilon_{\mathbf{v}}^2(\omega) &= E\{|\hat{V}(\omega) - \hat{V}_{\mathrm{N}}(\omega)|^2\}, \\ \varepsilon_{\mathbf{s}}^2(\omega) &= E\{|\hat{S}_{\mathrm{N}}(\omega)|^2\}. \end{aligned} \tag{10}$$

Optimization problem for the GSC method can be reformulated as follows:

$$\min_{\mathbf{g}(\omega)} \varepsilon_{\mathbf{v}}^2(\omega), \text{ subject to: } \varepsilon_{\mathbf{s}}^2(\omega) = \alpha(\omega), \tag{11}$$

where $\alpha(\omega)$ is a some predefined level of the speech leakage power. The complex Lagrange functional is given by:

$$L(\mathbf{g}(\omega), \lambda(\omega)) = \varepsilon_{\mathbf{v}}^2(\omega) + \lambda(\omega)(\varepsilon_{\mathbf{s}}^2(\omega) - \alpha(\omega)). \tag{12}$$

Differentiating (12) with respect to $\mathbf{g}(\omega)$ and equating to zero we find the solution:

$$\mathbf{g}_{\mathrm{SLC}}(\omega) = \mathbf{M}(\omega)^{-1}\mathbf{B}^H(\omega)\mathbf{R}_{\mathbf{vv}}(\omega)\mathbf{w}(\omega), \tag{13}$$

where

$$\mathbf{M}(\omega) = \mathbf{B}^H(\omega)[\mathbf{R}_{\mathbf{vv}}(\omega) + \lambda(\omega)\mathbf{R}_{\mathbf{yy}}(\omega)]\mathbf{B}(\omega). \tag{14}$$

The Lagrange multiplier $\lambda(\omega)$ provides a trade-off between speech leakage and noise reduction. It can be easily verified that for $\lambda(\omega) \to \infty$ speech leakage power is decreased at the expense of increased residual noise. If $\lambda(\omega) = 0$, the conventional GSC method is obtained.

The simplest approach is to set this parameter to empirically chosen fixed value. However an optimal (from the perceptual point of view) solution is to find λ_{opt} such that the speech distortion is inaudible and the residual noise is as low as possible. It can be done by substituting the masking threshold of the clean speech - $\phi_{\mathrm{m}}(\omega)$ for $\alpha(\omega)$ and solving the optimization constraint (11), i.e.:

$$\mathbf{g}_{\mathrm{SLC}}^H(\omega)\mathbf{B}^H(\omega)\mathbf{R}_{\mathbf{yy}}(\omega)\mathbf{B}(\omega)\mathbf{g}_{\mathrm{SLC}}(\omega) = \phi_{\mathrm{m}}(\omega), \tag{15}$$

In this way the speech distortions can be effectively reduced. This situation is also depicted in the Fig. 3. Unfortunately derivation of an explicit expression for $\lambda(\omega)$ seems to be a difficult task. It can be done numerically but we found that for certain cases the solution may not exists or be unstable (i.e. when the masking threshold level is very small). Therefore instead trying to solve (15) explicitly, we propose a suboptimal solution:

$$\lambda(\omega) = \lambda_{\max}\min(\mathrm{MNR}(\omega), 1), \tag{16}$$

where

$$\mathrm{MNR}(\omega) = \frac{\phi_{\mathrm{m}}(\omega)}{E\{|\hat{V}(\omega)|^2\}} = \frac{\phi_{\mathrm{m}}(\omega)}{\mathbf{w}^H(\omega)\mathbf{R}_{\mathbf{vv}}(\omega)\mathbf{w}(\omega)} \tag{17}$$

is the mask to noise ratio and $\lambda_{\max}$ is a maximum value for $\lambda(\omega)$. In our experiments it was empirically set to 0.25. The speech correlation matrix $\mathbf{R}_{\mathbf{yy}}(\omega)$ may be

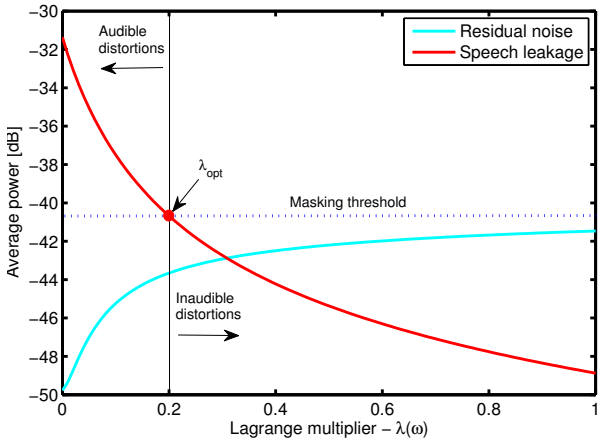

Fig. 3. Example of speech leakage masking

semi-positive definite thus the limiting the Lagrange multiplier improves a numerical stability of the matrix inversion in (13). Note that if the noise power level at beamformer output is below the masking threshold ($\mathrm{MNR}(\omega) \geq 1$) the noise is not audible, thus there is no need for noise cancellation and the speech leakage may be minimized as much as possible. Otherwise, if $0 < \mathrm{MNR}(\omega) < 1$, the noise is audible, thus $\lambda(\omega)$ is scaled proportionally to the MNR value, giving a better noise attenuation.

Theoretically instead of using the MNR one can use local signal-to-noise ratio (SNR), however it is known that the SNR estimate is rather erroneous and says nothing about masking effects [14]. In fact, most psychoacoustic models compute $\phi_{\mathrm{m}}(\omega)$ by performing some smoothing operations on speech power spectrum. Therefore we estimate the clean speech power spectral density (PSD), first:

$$\phi_{\mathrm{s}}(\omega) \approx E\{|\hat{S}(\omega)|^2\} = \mathbf{w}^H(\omega)\mathbf{R}_{\mathbf{yy}}(\omega)\mathbf{w}(\omega). \tag{18}$$

Then we use (18) as an input for Johnston's psychoacoustic model [9]. The correlation matrix of the microphone speech signal is computed as $\mathbf{R}_{\mathbf{yy}}(\omega) = \mathbf{R}_{\mathbf{xx}}(\omega) - \mathbf{R}_{\mathbf{vv}}(\omega)$.

4. Experiments

In this section we compare the performance of the conventional GSC beamformer with the proposed speech leakage constrained approach (denoted as GSC-SLC). The

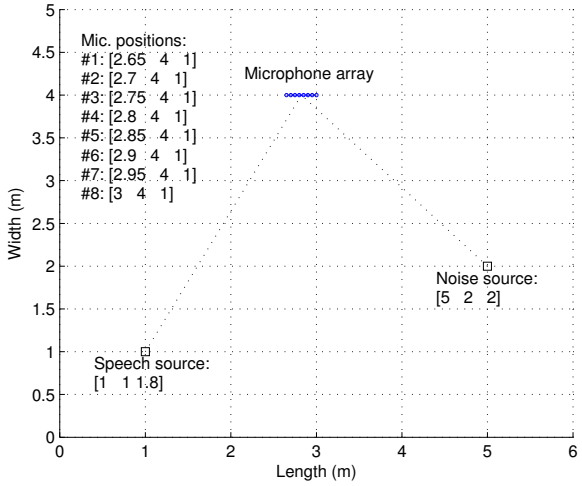

Fig. 4. Floor plan of the simulated enclosure (all coordinates in meters).

methods were implemented in MATLAB using overlap-save procedure. The microphone signals are cut into 50% overlapping frames of size $M = 1024$ samples that corresponds to time window of 128ms long (assuming 8kHz sampling rate). Once the signals are filtered in the DFT domain they are transformed back to time domain and only last $M/2$ samples are saved. In order to determine the system performance under model uncertainties we assumed simple direct-path acoustic model and use delay-and-sum beamformer, thus the steering vector and blocking matrix were computed using propagation delays. To efficiently compute the frequency filters the correlation matrix of the noise signal have to be estimated. However for comparative purposes we put aside this problem and compute $\mathbf{R}_{\mathbf{vv}}(\omega)$ directly from data. In practice any voice activity detector (VAD) can be used to update noise statistics in speech pauses only. Similarly we estimate microphone delays for delay-and-sum beamformer using an exact value of the direction of arrival (DOA) angle.

Two acoustic environments were simulated using the image method [1]: the first one with absorptive surfaces ($T_{60} = 33$ms) and the second one with reflective surfaces ($T_{60} = 135$ms). The parameter T_{60} denote reverberation time defined as the time taken for the sound to decay to 60dB below its value at cessation [10]. In both cases we assumed the rectangular enclosure with dimensions $6 \times 5 \times 2.8$ (all dimensions and coordinates are in meters). We considered an uniform linear array of 8 microphones placed on the x-axis with the first microphone at the position $(2.65, 4, 1)$ and spacing 0.05. The speech source signal was positioned at $(1, 1, 1.8)$. It was about 30s-long

comprised of eight shorter phonetically balanced sentences, uttered by eight of the speakers (four males and four females). These sentences have been selected from TIMIT database [6]. Originally they were recorded at 16kHz sampling rate but for our purposes they were low-pass filtered and downsampled to 8kHz. The noise source was located at $(5,2,2)$. The locations of the microphones and the sound sources are also depicted in Fig. 4. During the experiments two noise types have been considered: white Gaussian noise and babble noise both selected from NOISEX-92 database [13]. The microphone signals were obtained by convolving the speech source signal with the room impulse responses and adding to the corresponding noise signals at different SNRs, according to (1).

Our experiments were based on objective performance measurement. The amount of noise reduction was measured using noise attenuation factor defined as the mean ratio between the input noise power and output noise power. The speech distortion factor was defined as segmental signal to noise ratio where the noise is interpreted as a difference between the original and enhanced speech, thus the higher the factor the better. These measures mainly reflect the statistical differences between the signals. Therefore the cepstral distance [15] and modified Bark spectral distortion (MBSD) measure [16] were also used for evaluation of the audible differences. For computation of the cepstral distance we use first 16 cepstral coefficients, that are excepted to carry tonal information. The lower the cepstral distance and/or MBSD measure, the less audible speech distortion. Additionally PESQ measure [8] was exploited for overall evaluation of the speech quality.

The objective measurement results are depicted in Fig. 5. The vertical error-bars denote 95% confidence intervals estimated using 1000 bootstrap data samples. As can be seen the relative improvements are similar for both noise types. Note that the noise statistics were estimated directly from data, often noise signal is not directly available and noise statistics must be estimated from the noisy speech signal, i.e. during speech pauses, thus in practice some performance drop is expected.

For non-reverberant environment ($T_{60} = 33$ms) the improvement is rather not significant. It is not surprising since in this case the direct path model is accurate enough (speech signal goes straight from the sound source to the listener), thus speech leakage is very low and the parameter $\lambda(\omega)$ has no impact on the system performance. In this case the proposed method is equivalent to the conventional GSC beamformer.

In the case of reverberant environment ($T_{60} = 135$ms) the direct path model is not sufficient (i.e. presence of the system model uncertainties) which results in increased speech leakage. However as can be seen in Fig. 5 (solid lines) the proposed method outperforms conventional one providing significantly better performance at lower SNRs in the terms of speech distortion and MBSD measure. One exception

is cepstral distance where confidence intervals are slightly overlapped and thus the improvement is not significant. In order to avoid overestimation of the noise attenuation factor, it should be measured in speech pauses only, however it is rather difficult to precisely mark these regions. Thus, this factor was estimated also in transients where mean squared error is substantially lower for the speech leakage constrained method. Theoretically this measure should be comparable for both methods. On the

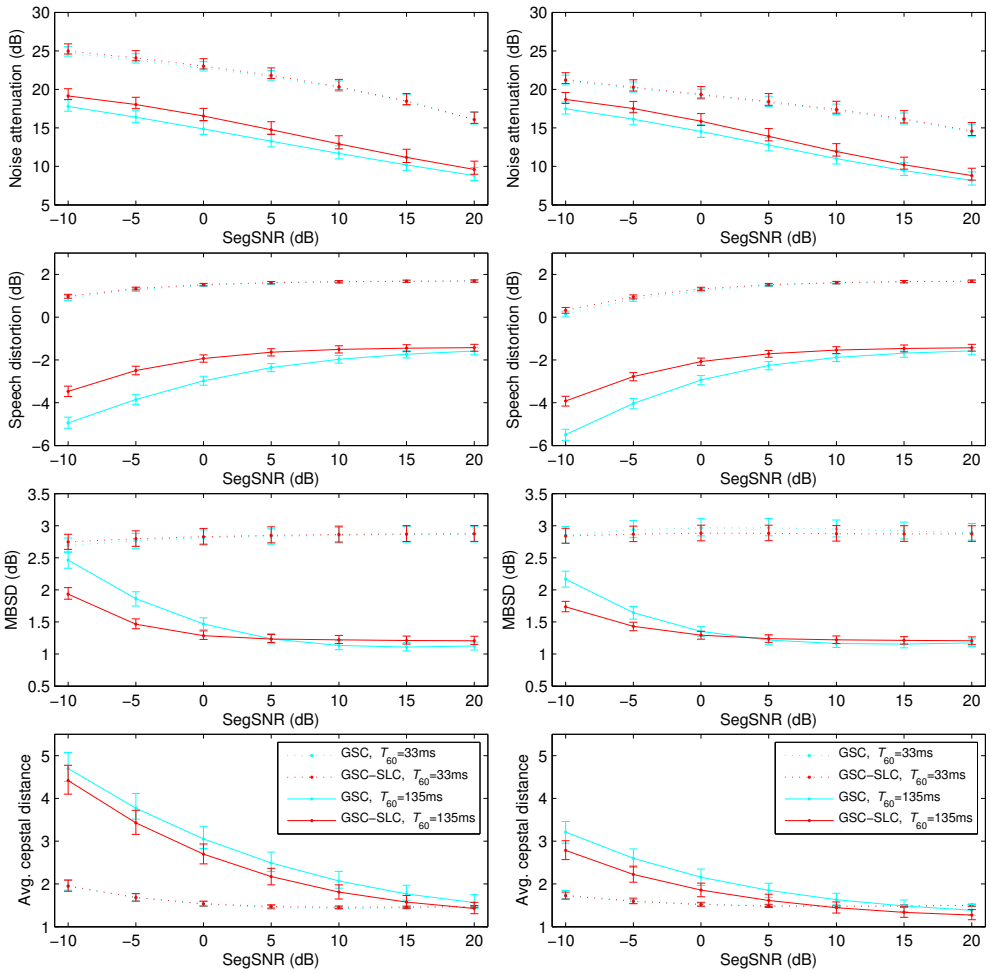

Fig. 5. Comparison of the objective performance measures for the conventional GSC beamformer and the proposed GSC-SLC method in two noisy environments: white noise (left) and babble noise (right); vertical lines denote 95% confidence intervals.

Table 1. Perceptual evaluation using PESQ.

	White Noise				Babble Noise			
	$T_{60} = 33$ms		$T_{60} = 135$ms		$T_{60} = 33$ms		$T_{60} = 135$ms	
SegSNR	GSC	GSC-SLC	GSC	GSC-SLC	GSC	GSC-SLC	GSC	GSC-SLC
-10	2.350	2.412	1.797	1.800	2.306	2.429	1.916	1.988
-5	2.575	2.648	1.914	1.992	2.548	2.643	2.027	2.147
0	2.775	2.847	2.041	2.139	2.757	2.838	2.133	2.262
5	2.947	3.006	2.150	2.262	2.945	3.021	2.243	2.340
10	3.113	3.165	2.249	2.343	3.128	3.191	2.335	2.389
15	3.293	3.356	2.334	2.387	3.301	3.369	2.390	2.437
20	3.479	3.566	2.404	2.423	3.468	3.542	2.433	2.454

other hand, it is clear that a residual noise increase is not proportional to the speech distortion decrease. In our experiments this increase is 'negative'.

Similar observations can be made for the PESQ scores (see Tab. 1). Although we observe lower performance results for the conventional GSC beamformer for both reverberation/noise conditions, in the case of reverberant environment relative improvement is higher.

5. Conclusion

The performance of the conventional GSC beamformer can be improved in the presence system model uncertainties by using auditory properties. We derived a noise cancellation filter which is able to reduce the speech leakage (and speech distortions) at expense of residual noise increase. However as we show this increase is rather small. In addition it is tolerated by auditory system as long as the noise level is placed below masking threshold. The experimental results show that the proposed method outperforms conventional GSC beamformer providing lower speech distortions and comparable residual noise level.

There are some possible improvements of the proposed method, i.e.: a derivation of an explicit formula for optimal Lagrange multiplier, a recursive implementation of the frequency filters or an estimation of the steering vector and blocking matrix using second-order statistics only. These issues will be considered in a future work.

Acknowledgment

This work was supported by Bialystok University of Technology under the grant S/WI/1/2013.

References

[1] J.B. Allen and D.A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal Acoustic Society of America*, 65(4):943, 1979.

[2] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski. On microphonearray beamforming from a mimo acoustic signal processing perspective. *IEEE Trans. Audio, Speech, Lang. Process.*, 15(3):1053–1065, 2007.

[3] A. Borowicz and A. Petrovsky. Signal subspace approach for psychoacoustically motivated speech enhancement. *Speech Comm.*, 53(2):210–219, 2011.

[4] O.L. Frost. An algorithm for linearly constrained adaptive array processing. In *Proc. IEEE*, volume 60, pages 926–935, Aug 1972.

[5] S. Gannot, D. Burshtein, and E. Winstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Trans. Signal Process.*, 49(8):1614–1626, 2001.

[6] J.S. Garofolo, L.F. Lamel, W.M. Fisher, J.G. Fiscus, D.S. Pallett, and N.L. Dahlgren. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. National Institute of Standards and Technology (NIST), CD-ROM, 1993.

[7] Y. Huang, J. Benesty, and J. Chen. Analysis and comparison of multichannel noise reduction methods in a common framework. *IEEE Trans. Audio, Speech, Lang. Process.*, 16(5):957–968, 2008.

[8] ITU-T. Perceptual evaluation of speech quality (PESQ). Rec. P.862, ITU, Geneva, 2001.

[9] J.D. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE J. on Selected Areas in Comm.*, 6:314–323, February 1988.

[10] R. Ratnam, D.L. Jones, and W.D. O'Brien. Fast algorithms for blind estimation of reverberation time. *IEEE Signal Process. Lett.*, 11(6):537–540, 2004.

[11] M. Souden, J. Benesty, and S. Affes. A study of the LCMV and MVDR noise reduction filters. *IEEE Trans. Sig. Process.*, 58(9):4925–4935, Sept 2010.

[12] R. Talmon, I. Cohen, and S. Gannot. Convolutive transfer function generalized sidelobe canceler. *IEEE Trans. Speech, Lang. Process.*, 17(7):1420–1434, Sept 2009.

[13] A. Varga and H.J.M. Steeneken. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication*, 12(3):247–251, 1993.

[14] D. Virette, P. Scalart, and C. Lamblin. Analysis of background noise reduction techniques for robust speech coding. In *Proc. EUSIPCO*, volume 3, pages 297–300, 2002.

[15] S. Wang, A. Sekey, and A. Gersho. An objective measure for predicting subjective quality of speech coders. *IEEE J. Sel. Areas Commun.*, 10:819–829, 1992.

[16] W. Yang, M. Benbouchta, and R. Yantorno. Performance of a modified bark spectral distortion measure as an objective speech quality measure. In *Proc. ICASSP*, pages 541–544, Seattle, USA, 1998.

SKUTECZNY TŁUMIK LISTKÓW BOCZNYCH Z WYKORZYSTANIEM MASKOWANIA PRZECIEKU MOWY

Streszczenie: Prezentowana jest nowa metoda uzdatniania mowy w oparciu o strukturę uogólnionego tłumika listków bocznych. Wykazujemy, że możliwe jest zmniejszenie słyszalnych zniekształceń mowy przy zachowaniu stałego poziomu szumu rezydualnego, dla modeli przybliżonych środowiska akustycznego. Może to być dokonane poprzez uwarunkowanie poziomu mocy przecieku mowy zgodnie ze zjawiskiem maskowania oraz minimalizację warunkową mocy szumu rezydualnego. Proponowane podejście zaimplementowano w oparciu o prosty model beamformera opóźniająco-sumującego. Ostatecznie przeprowadzono ocenę porównawczą wybranych metod z wykorzystaniem obiektywnych miar jakości mowy. Wyniki pokazują, że nowa metoda przewyższa konwencjonalną zapewniając mniejsze zniekształcenia mowy.

Słowa kluczowe: GSC, psychoakustyka, uzdatnianie mowy

Artykuł zrealizowano w ramach pracy badawczej S/WI/1/2013.