EVALUATION OF THE PENALIZED INEQUALITY CONSTRAINED MINIMUM VARIANCE BEAMFORMER FOR HEARING AIDS

Jinjun Xiao[1], *Wenqiang Pu*[2,3], *Zhi-Quan Luo*[3], *Tao Zhang*[1]

[1] Starkey Hearing Technologies, Eden Prairie, MN 55344, USA
[2] National Lab of Radar Signal Processing, Xidian University, Xi'an, China
[3] Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China
{jinjun_xiao,tao_zhang}@starkey.com, wenqiangpu@outlook.com, luozq@cuhk.edu.cn

ABSTRACT

Beamforming is a common technique used to improve speech intelligibility and listening comfort of hearing aids users in a noisy environment. Traditional beamforming algorithms such as linearly constrained minimum variance (LCMV) beamformer cannot effectively suppress multiple interferences when the degree of freedom (DoF) of the array is less than the number of sources in the environment. In [1], a penalized inequality-constrained minimum variance (P-ICMV) beamformer was proposed to address this challenge. In this study, we evaluate the P-ICMV beamformer and compare its performance with other beamformers including the LCMV in a multiple-interference environment. In an objective evaluation, objective metrics related to speech intelligibility and sound quality are used to compare the algorithm performance. In a subjective evaluation, the speech intelligibility of the beamformer processed stimuli are evaluated using normal-hearing listeners. Both the objective and subjective evaluation results show that the P-ICMV beamformer can suppress the interferences more effectively than the existing beamformers when the array DoF is limited.

Index Terms**—** Microphone array signal processing, acoustic beamforming, binaural signal processing, speech intelligibility

1. INTRODUCTION

Various research results have indicated that hearing-impaired listeners require a better signal-to-noise ratio (SNR) than normal-hearing listeners to understand speech in noise [2, 3]. Modern hearing aids equipped with microphone array leveraging beamforming techniques have been proven to effectively improve the listening experience of the hearing aids users [4, 5]. Traditional adaptive beamforming algorithms such as linearly constrained minimum variance (LCMV) beamformer are not robust against factors such as uncertainty in the array response [6, 7, 8]. Such algorithms, when applied on hearing aids, cannot effectively suppress multiple interferences in a complex listening environment due to the small number of microphones in the array.

The traditional LCMV beamformer assumes perfect knowledge of the relative transfer functions (RTFs) from sources to microphones. The RTFs are used in linear equality constraints for target protection and interference suppression. In a hearing aids application, the real-time estimation of the RTFs is difficult due to factors such as the wearer's head movement and the change of hearing aids placement on the ear. Hence robust beamformer algorithms [9] are desired. To handle the direction of arrival (DoA) estimation error, a robust beamformer called inequality constrained minimum variance (ICMV) beamformer was proposed in [10]. The ICMV adopts the approach of relaxing equality constraints to a number of inequalities for robust target protection.

In both LCMV and ICMV, the number of interferences that can be handled by the beamformer is limited by the DoF offered by the array. Such limitation restricts their application in certain multiple-interference environments. In [1], a penalized-ICMV (or P-ICMV) beamformer was proposed to address this challenge. The P-ICMV beamformer imposes inequality constraints with a maximum gain on each interference. The maximum gain is penalized in the optimization objective. This makes the formulation feasible even when the number of constraints is larger than the array DoF, and thus allows the beamformer to handle any number of interferences.

In this study, we evaluate the P-ICMV beamformer and compare its performance with the LCMV and ICMV beamformers in a multiple-interference environment. We compare the signal to noise ratio improvement and target distortion of these three beamformers under a couple of setups. The hearing-aid speech perception index (HASPI) and Hearing Aid Speech Quality Index (HASQI) [11, 12] are used to evaluate the speech intelligibility and sound quality of the beamformer outputs objectively. In addition, a subjective Evaluation is used to confirm the speech intelligibility prediction based on obtained HASPI scores.

The remainder of this paper is organized as follows. Section 2 summarizes the LCMV, ICMV and P-ICMV beamformers. Section 3 presents the objective and subjective evaluation results. Section 4 gives the conclusion of the paper.

2. THE P-ICMV BEAMFORMER

Let us consider the beamforming problem on a pair of binaural hearing aids with M microphones on each side. We assume that the environment has one target source and K interfering sources. The received signals at the microphones in the time-frequency domain can be expressed as

$$\mathbf{y}(l,f) = \mathbf{x}(l,f) + \mathbf{v}(l,f) + \mathbf{n}(l,f)$$

where

$$\mathbf{x}(l,f) = \mathbf{h}_{\mathrm{s}}(f)s(l,f); \;\; \mathbf{v}(l,f) = \sum_{k=1}^{K} \mathbf{h}_k(f) i_k(l,f).$$

In the above equations, (l,f) are the time and frequency indices; $\mathbf{h}$ are the ATFs; s and i are the target and interference signals respectively; and $\mathbf{n}$ is the background noise. To simplify the notation, the indices l and f are omitted in the rest of the paper.

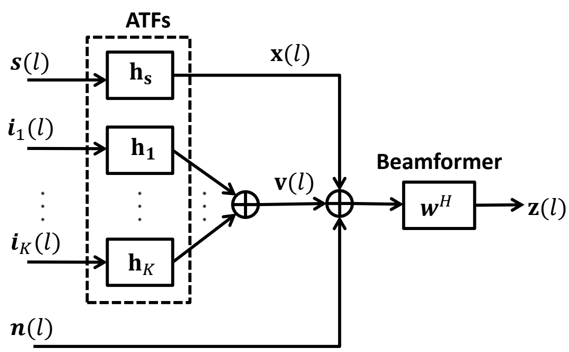


Fig. 1. The schematic of the considered system.

The hearing aids apply a beamformer $\mathbf{w}$ to the received signal $\mathbf{y}$ and linearly combine the filtered signals to produce the output signal, z, as (see Fig. 1)

$$z = \mathbf{w}^H\mathbf{y} = \underbrace{\mathbf{w}^H\mathbf{x}}_{\text{desired speech}} + \underbrace{\mathbf{w}^H\mathbf{v}}_{\text{interferences}} + \underbrace{\mathbf{w}^H\mathbf{n}}_{\text{noise}}.$$

A beamformer is designed to balance among noise reduction, interference suppression, and target protection. With different designs, various beamformers have been proposed. With the *a priori* knowledge of RTFs (with respect to the ref. mic): $\bar{\mathbf{h}}_\theta$, the LCMV beamformer minimizes the noise residual subject to equality constraints of suppressing the interferences and protecting the target:

$$\min_{\mathbf{w}} \quad \mathbf{w}^H\mathbf{R}_n\mathbf{w} \tag{1a}$$
$$\text{s.t.} \quad \bar{\mathbf{h}}_{\mathrm{s}}^H\mathbf{w} = 1 \tag{1b}$$
$$\bar{\mathbf{h}}_k^H\mathbf{w} = 0, \; k = 1,\dots,K \tag{1c}$$

where $\mathbf{R}_n \triangleq \mathbb{E}[\mathbf{n}\mathbf{n}^H]$ is the noise correlation matrix.

To improve the robustness of the LCMV beamformer against the RTF estimation errors, a robust version of the LCMV was proposed by changing the equality constraints to inequality constraints [10], which results in the so-called ICMV beamformer:

$$\min_{\mathbf{w}} \quad \mathbf{w}^H\mathbf{R}_n\mathbf{w} \tag{2a}$$
$$\text{s.t.} \quad |\bar{\mathbf{h}}_\theta^H\mathbf{w} - 1|^2 \le c_\theta^2, \; \forall\theta \in \Theta \tag{2b}$$
$$|\bar{\mathbf{h}}_\phi^H\mathbf{w}|^2 \le c_\phi^2, \; \forall\phi \in \Phi_k, \; k = 1,\dots,K. \tag{2c}$$

The discrete angle set Θ is pre-specified as a desired angle aperture around the target to handle the DoA errors, e.g., $\Theta = \{\eta - 10^\circ, \eta, \eta + 10^\circ\}$, where η is the estimated DoA. The constant c_θ provides a user selected tolerable speech distortion threshold at $\theta \in \Theta$. The notations $\bar{\mathbf{h}}_\theta$ and $\bar{\mathbf{h}}_\phi$ denote the RTFs at angle θ and ϕ respectively. The parameter c_ϕ^2 specifies an upper bound on the interference suppression. Similarly, to improve the robustness of interference suppression, multiple constraints for angles within a set Φ_k around the estimated interference DoA can be enforced.

One limitation of the LCMV and ICMV beamformers is their ability of handling multiple interferences. When the DoFs of the array are fewer than the number of sources in the environment (i.e., $2M \le K$), both beamformers become infeasible. Thus, in such a case, only part of interference sources can be effectively suppressed. The P-ICMV beamformer proposed in [1] extends the ICMV into a penalized version. Compared to ICMV, P-ICMV has an extra optimization variable $\boldsymbol{\epsilon}$ which makes the upper bound on $|\bar{\mathbf{h}}_\phi^H\mathbf{w}|^2$ part of the optimization problem (instead of being pre-determined):

$$\min_{\mathbf{w},\boldsymbol{\epsilon}} \quad \mathbf{w}^H\mathbf{R}_n\mathbf{w} + \mu \max_k\{\gamma_k\epsilon_k\} \tag{3a}$$
$$\text{s.t.} \quad |\bar{\mathbf{h}}_\theta^H\mathbf{w} - 1|^2 \le c_\theta^2, \; \forall\theta \in \Theta \tag{3b}$$
$$|\bar{\mathbf{h}}_\phi^H\mathbf{w}|^2 \le \epsilon_k c_\phi^2, \; \forall\phi \in \Phi_k, \; k = 1,\dots,K. \tag{3c}$$

We can see that the number of constraints for interference suppression is no longer limited by the DoF. Thus P-ICMV is able to produce a feasible solution for arbitrary number of interferences so long as $2M \ge |\Theta|$. An additional parameter μ is introduced in the objective function to provide a tradeoff between the noise reduction and interference suppression.

The optimization problem of the ICMV and P-ICMV beamformers is a second-order cone programming. Efficient optimization algorithm based on the alternating direction method of multipliers method has been derived and successfully applied to solve the problem in [10, 1] respectively.

3. EVALUATION

In this section, we present the objective and subjective evaluation results of the three beamformers.

3.1. Acoustic Conditions

A simulated room of size 12.7m×10m with height 3.6m is used for the evaluation (Fig. 2). The room reverberation time is chosen to be 0.6 second. The hearing aids wearer is located at the center of the room. The target and interference sources are represented by speakers 1.0m away from the listener. The background babble noise is simulated using 24 speakers at different locations illustrated in the figure. All speakers and hearings aids microphones are in the same horizontal plane at a height of 1.2m. The room impulse responses (RIRs) is generated by the so-called image method [13].

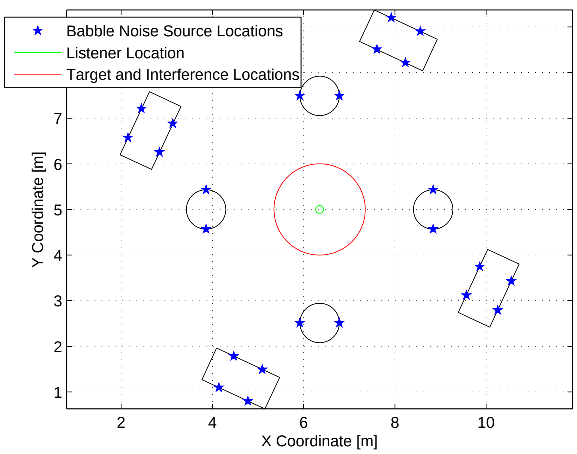
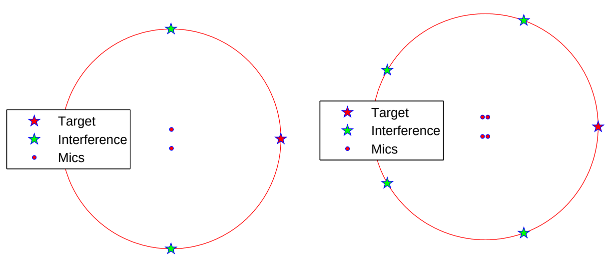


Fig. 2. Simulated acoustic environment.



(a) Setup 1 (b) Setup 2

Fig. 3. Target and interference locations

There are two acoustic setups used in the evaluation. In both setups, the target is at 0°. The background noise is set to a level so that the SNR at the left reference microphone is 10dB. The numbers of interferences and microphones are different in these two setups. In setup 1, each hearing aid has 1 microphone. There are 2 interferences located at ±90° respectively (see left plot in Fig. 3). The two interferences have the same level and are chosen at 3dB lower than the target speech, so the resulted SIR is 0dB. In setup 2, each hearing aid has 2 microphones with 7.5mm spacing. The front microphone is set as the reference mic. There are 4 interferences located at ±70° and ±150° respectively (see right plot in Fig. 3). The two front interferences are set to be 1dB higher than the target speech and the two rear interferences are set to be 4dB lower than the target speech. The resulted SIR is about −5dB.

In both setups, the number of microphones in the array is fewer than the total number of target and interference sources in the environment. Our goal is to evaluate how the P-ICMV handles the extra source in such conditions. In the simulations, we assume that the DoA of each source is known. The microphone signals are generated using the reverberant ATF. In addition, the anechoic ATF corresponding to the given source DoA is used in the beamformer calculation.

3.2. Objective Evaluation

The intelligibility-weighted SINR improvement (IW-SINRI) and IW-spectral distortion (IW-SD) have been calculated and compared in [1]. For the completeness of the paper we list the IW-SINRI and IW-SD scores of these two setups in Table 1. Due to the insufficient number of microphones, the LCMV and ICMV have to ignore one interference in each setup. In setup 1, the two interferences have equal levels and the one at 90° is ignored. In setup 2, the weaker interference at −70° is ignored.

Table 1. IW-SINRI and IW-SD[dB]

Setup	LCMV 1	LCMV 2	ICMV 1	ICMV 2	P-ICMV 1	P-ICMV 2
IW-SINRI	0.62	2.31	0.65	2.48	**3.38**	**9.22**
IW-SD	**0.42**	2.05	0.51	2.06	0.50	**1.27**

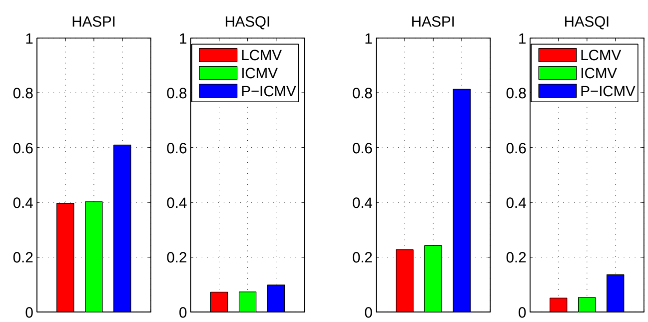


Fig. 4. The objective perceptual scores of setup 1 (left) and setup 2 (right).

Objective perception metrics HASPI and HASQI [11, 12] have been proposed to measure speech intelligibility and sound quality respectively for both normal-hearing and hearing-impaired listeners. These metrics have been shown to be closely correlated with the subjective test results [14]. We thus use them to evaluate the speech intelligibility and sound quality of the three beamformer outputs. The scores are given in Fig. 4. Both scores have a range from 0 to 1. The better speech intelligibility improvement of the P-ICMV beamformer over the other two beamformers is clearly shown in the HASPI scores. The amount of HASPI improvement of P-ICMV over LCMV and ICMV in setup 2 is greater than that in setup 1. This is consistent with the IW-SINRI result above. The HASQI scores indicate that the sound quality among the three beamformer outputs is similar.

3.3. Subjective Evaluation

In this section, we describe a subjective test performed to verify the speech intelligibility predicted by the HASPI scores in previous section[1]. In total 12 subjects (10 males and 2 females) participated in this study. All subjects have normal hearing and normal cognitive functions. They were all employees of Starkey Hearing Techologies, Inc. and were not paid for their participation in this study.

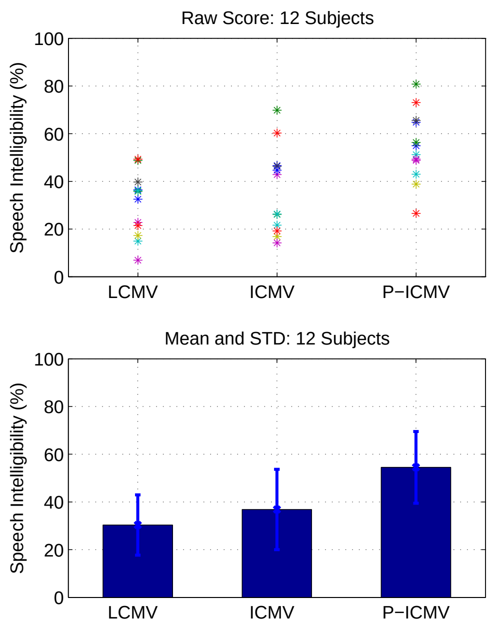


Fig. 5. Subjective evaluation scores (setup 1).

The stimuli were created as follows. The target speech is from the Connected Speech Test (CST) speech corpus [15] and interferences are from the International Speech Test Signal (ISTS) [16]. The background babble noise was simulated using the 24 speaker locations described in Fig. 2. All three beamformer outputs under the two setups described in previous section are evaluated.

For each subject, 12 passages from the CST speech corpus were used as target speech (2 passages per beamformer per setup). Besides the 12 testing paragraphs, each subject had 2 additional passages for practice to get acquainted with the test. Each passage has 10 sentences and 25 key words used to score the intelligibility. Each passage is about a specific topic (such as "lawn" or "cactus") that was known to the subject during the test. The stimuli were played at 7dBSPL through a Beyer Dynamic DT770 Pro headphones in a sound treated room. Subjects were asked to repeat the sentence heard, which was then compared to the scripts with key words scored to obtain the intelligibility percentage.

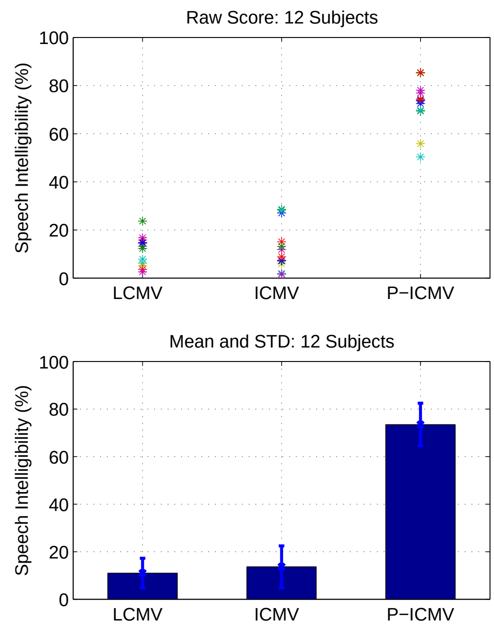


Fig. 6. Subjective evaluation scores (setup 2).

The top panels in Fig. 5 and 6 show the distribution of the raw scores while the bottom panels show the mean and standard deviation. It can be seen that the relative speech intelligibility improvement among the three beamformers is close to what is predicted by the HASPI scores in previous section, while the overall score from the subjective test is slightly lower than the HASPI scores. Another observation is the large variation among subjects (see the raw score distribution in the top panels in Fig. 5 and 6). This variation can be explained by the different noise tolerance by the subjects and also the overall effort each subject made to understand the target speech in a very noisy environment.

4. CONCLUSION

In this paper, we evaluated the P-ICMV beamforming algorithm by comparing its performance with the LCMV and ICMV beamformers. The P-ICMV beamformer's ability to effectively suppress a larger number of interferences than the number of microphones is demonstrated through both objective and subjective evaluations. Results show that the P-ICMV beamformer offers significantly improved interference suppression and better speech intelligibility improvement while maintaining similar sound quality in a setup where the DoF is limited.

[1]Due to the lack of time, the speech quality subjective evaluation is planned as part of the future research.

5. REFERENCES

[1] W. Pu, J. Xiao, T. Zhang, and Z.-Q. Luo, "A penalized inequality-constrained minimum variance beamformer with applications in hearing aids," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2017.

[2] S. Kochkin, "Customer satisfaction and subjective benefit with high performance hearing aids," *Hearing Review*, vol. 3, no. 12, pp. 16–26, 1996.

[3] M. Killion, "SNR loss: I can hear what people say, but i can't understand them," *Hearing Review*, vol. 4, no. 12, pp. 8–14, 1996.

[4] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1365–1376, October 1987.

[5] R. A. Bentler, C. Palmer, and A. B. Dittberner, "Hearing-in-noise: Comparison of listeners with normal and (aided) impaired hearing," *Journal of the American Academy of Audiology*, vol. 15, pp. 216–225, 2004.

[6] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1662–1676, March 1992.

[7] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, March 2015.

[8] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 543–558, March 2016.

[9] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 313–324, February 2003.

[10] W. C. Liao, M. Hong, I. Merks, T. Zhang, and Z. Q. Luo, "Incorporating spatial information in binaural beamforming for noise suppression in hearing aids," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 5733–5737.

[11] J. Kates and K. Arehart, "The hearing-aid speech quality index (HASQI)," *Journal of the Audio Engineering Society*, vol. 58, no. 5, pp. 363–381, 2010.

[12] J. Kates and K. Arehart, "The hearing-aid speech quality index (HASQI) version 2," *Journal of the Audio Engineering Society*, vol. 62, no. 3, pp. 99–117, 2014.

[13] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[14] A. A. Kressner, D. V. Anderson, and C. J. Rozell, "Robustness of the hearing aid speech quality index (HASQI)," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2011.

[15] R. M. Cox, G. C. Alexander, and C. Gilmore, "Development of the connected speech test (CST)," *Ear Hear*, p. 8, October 1987.

[16] I. Holube, M. Vlaming S. Fredelake, and B. Kollmeier, "Development and analysis of an international speech test signal (ISTS)," *International Journal of Audiology*, vol. 49, pp. 891–903, 2010.