# A New Design Method for Broadband Microphone Arrays for Speech Input in Automobiles

Ka Fai Cedric Yiu, Nedelko Grbić, *Associate Member, IEEE*, Kok-Lay Teo, *Senior Member, IEEE*, and Sven Nordholm, *Member, IEEE*

***Abstract*—In this letter, a new design method for broadband microphone arrays is presented. Using sequences of calibration signals, the method is able to design finite-impluse response (FIR) filters with specific performance. The method can control and adjust the speech distortion, noise suppression, and echo cancellation directly. It turns out that significantly shorter filter length can be applied to achieve better overall performance than the least-squares method or the signal-to-noise plus interference method.**

***Index Terms*—Microphone array, nonlinear programming, Pareto optimum, speech enhancement.**

## I. Introduction

THE INCREASED popularity of wireless cellular telephones and their uses in cars has motivated the development of hands-free in-car communication devices. In this particular acoustic environment, the microphone array is required to suppress the car noise as well as the echo from the hands-free loudspeaker while leaving the distortion of the speech to a minimum. Because this problem is very difficult to be described by *a priori* models, sequences of calibration signals are often used for the design of the beamformer [2].

In general, the optimal beamformer design problem is a multicriteria decision problem, where the criteria are the level of distortion, the level of noise suppression, and the level of interference suppression. In present designs, the least-squares technique (LS) and the signal-to-noise plus interference ratio (SNIR) are often used [1], [3] to optimize for the performance of the beamformer. However, in both approaches, there is no direct control over individual criteria. During the design process, it is not possible to specify the performance of the beamformer in advance. Furthermore, once the optimal weights are found, if some of the criteria are not satisfactory, there is no direct method to adjust the weights for those criteria. Nevertheless, it is a yield-acceptable design for all criteria. Several authors have proposed to control the noise suppression and the distortion levels by means of adjusting a Langrange multiplier [4], [5]. However, echo suppression has generally been ignored, and it is not straightforward to extend the Langrange multiplier approach to control the echo suppression as well.

Here, we propose a new method which is based on the nonlinear programming technique for solving the multicriteria design for the beamformer. This approach is versatile in the way that the specific performance of the beamformer can be imposed in advance. The suppression and distortion levels can be controlled easily, and the corresponding optimal weights can be sought. It is shown that significantly shorter filter length can be used than the LS and SNIR methods, and better overall performance can be achieved.

## II. Formulation of the New Method

Assume there are $M$ elements in the microphone array. In general, the signals received by the microphone element can be represented by

$$x^i(n) = x_s^i(n) + x_N^i(n) + x_I^i(n), \quad i = 1, 2, \ldots, M \quad (1)$$

where $x_s^i(n)$, $x_N^i(n)$, and $x_I^i(n)$ are the source signal, the noise signal, and the echo signal, respectively. Assume that known calibration sequence observations are used for each of these signals. The output of the beamformer is given by

$$y(n) = \sum_{i=1}^{M} \boldsymbol{w}_i^T \boldsymbol{x}^i(n) \quad (2)$$

where the weight vectors and the input data vectors are both of length $L$, the length of the filters. Let $\boldsymbol{W} = (\boldsymbol{w}^1, \ldots, \boldsymbol{w}^M)$ be the matrix of weights; the least-squares formulation will try to seek the weight matrix $\boldsymbol{W}$ via

$$\boldsymbol{w}_{\text{opt}}^i = \arg\min_{\boldsymbol{w}^i} \left\{ \sum_{n=0}^{K-1} (y(n) - x_s^i(n))^2 \right\} \quad (3)$$

where $K$ is the number of sampling points and $i \in [1, \ldots, M]$. In the signal-to-noise plus interference formulation, the weight matrix is sought to maximize the signal-to-noise plus interference power ratio [2]. In both approaches, there is no direct control over the level of distortion and the level of noise and interference suppressions. Evaluation results using very long filters [3] have shown that beamformers designed by LS had very good distortion controls, but poor in the suppression levels, while the ones designed by SNIR had very good suppression controls but left consistently high distortion levels. Both approaches are poor in controlling the balance between the noise and the interference suppression.

The objectives for the beamformer are to maximize the noise and interference suppressions, while keeping distortion caused

Manuscript received January 15, 2001; revised October 25, 2001. K. F. C. Yiu and K.-L. Teo were supported by the Research Committee of the Hong Kong Polytechnic University. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shuichi Ohno.

K. F. C. Yiu and K. L. Teo are with the Department of Applied Mathematics and the Centre for Multimedia Signal Processing, Hong Kong Polytechnic University, Kowloon, Hong Kong.

N. Grbić is with the Department of Telecommunications and Signal Processing, University of Karlskrona/Ronneby, 372 25 Ronneby, Sweden.

S. Nordholm is with the Australian Telecommunications Research Institute, Curtin University of Technology, Perth, Australia.

Publisher Item Identifier 10.1109/LSP.2002.802109.

by beamforming filters at the minimum. In order to measure various quantities, calibration signals are used. Define $\hat{P}_{x_s}(\omega)$ as the spectral power estimate of the source signal without filtering, and $\hat{P}_{y_s}(\omega)$ is the spectral power estimate of the beamformer output with filtering when the source signal is active alone. The normalized distortion measure can be defined as

$$D(\boldsymbol{W}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |C_d \hat{P}_{y_s}(\omega) - \hat{P}_{x_s}(\omega)|\, d\omega \tag{4}$$

where $\omega = 2\pi f$ and $f$ is the normalized frequency. The constant $C_d$ is defined as

$$C_d = \frac{\int_{-\pi}^{\pi} \hat{P}_{x_s}(\omega)\, d\omega}{\int_{-\pi}^{\pi} \hat{P}_{y_s}(\omega)\, d\omega}. \tag{5}$$

Similarly, the normalized noise suppression measure and interference suppression measure can be introduced as

$$S_N(\boldsymbol{W}) = C_s \frac{\int_{-\pi}^{\pi} \hat{P}_{y_N}(\omega)\, d\omega}{\int_{-\pi}^{\pi} \hat{P}_{x_N}(\omega)\, d\omega} \tag{6}$$

and

$$S_I(\boldsymbol{W}) = C_s \frac{\int_{-\pi}^{\pi} \hat{P}_{y_I}(\omega)\, d\omega}{\int_{-\pi}^{\pi} \hat{P}_{x_I}(\omega)\, d\omega} \tag{7}$$

where $C_s = 1/C_d$. Note that in (6), $\hat{P}_{y_N}(\omega)$ and $\hat{P}_{x_N}(\omega)$ are spectral power estimates of the beamformer output with and without filtering, when the surrounding noise is active alone. In the same manner, $\hat{P}_{y_I}(\omega)$ and $\hat{P}_{x_I}(\omega)$ are spectral power estimates when the interence signals are active alone. Both the noise and the interference suppression measures are normalized to the amplification/attenuation caused by the beamformer to the reference observation when the source signal is active alone, i.e., when the beamformer attenuates the source signal by a specific amount, the noise and interference suppression quantities are reduced by the same amount.

Because there is more than one objective in the design of the beamformers, it is basically a multicriteria design problem [6], [7]. When different scaling factors are applied to the criteria in the design process, a solution set can be derived in which all solutions are efficient, or Pareto optimum. In the present context, the set of weights $\boldsymbol{W}^*$ is Pareto optimum if and only if there does not exist a set of weights $\boldsymbol{W}$ such that

$$S_N(\boldsymbol{W}) \geq S_N(\boldsymbol{W}^*) \qquad S_I(\boldsymbol{W}) \geq S_I(\boldsymbol{W}^*) \tag{8}$$

$$D(\boldsymbol{W}) \leq D(\boldsymbol{W}^*) \tag{9}$$

with strict inequality to at least one of the criteria. In order to solve for the Pareto optimum, some of the criteria can be formulated as constraints instead so that it becomes a nonlinear programming problem. An additional advantage of using this formulation is that the constraints can be adjusted freely to select the desired filter from the set of Pareto optimum solutions.

As a first step, in order to control the suppression and the distortion separately, the optimal design problem can be converted into an equivalent nonlinear programming problem (P1) as

$$\min_{\boldsymbol{W}} F_1(\boldsymbol{W}) \equiv -(S_N(\boldsymbol{W}) + S_I(\boldsymbol{W})) \tag{10}$$

subject to

$$D(\boldsymbol{W}) \leq d \tag{11}$$

TABLE I

| $L = 16$ | $D[dB]$ | $S_N[dB]$ | $S_I[dB]$ |
|---|---|---|---|
| LS | -28.79 | 15.95 | 9.24 |
| SNIB | -19.28 | 8.96 | 4.13 |
| New method | -25. | 15.24 | 15.19 |

TABLE II

| $L = 32$ | $D$ | $S_N$ | $S_I$ |
|---|---|---|---|
| LS | -29.14 | 16.38 | 11.01 |
| SNIB | -19.20 | 14.98 | 9.97 |
| New method | -25. | 17.82 | 17. |

where $d$ is a predefined distortion level. The level of distortion can now be controlled freely, but it is still a bicriteria design problem. To control the noise suppression and the interference suppression separately, one of the criteria, e.g., the suppression level, can be formulated as a constraint, and the corresponding equivalent nonlinear programming problem (P2) is

$$\min_{\boldsymbol{W}} F_2(\boldsymbol{W}) \equiv -S_N(\boldsymbol{W}) \tag{12}$$

subject to

$$S_I(\boldsymbol{W}) \geq s_i \tag{13}$$

$$D(\boldsymbol{W}) \leq d \tag{14}$$

where $s_i$ is a predefined interference suppression level. Note that the role of $S_N$ and $S_I$ can be exchanged in this formulation so that either one can be in the objective function and the other as a constraint. The only problem with this formulation is how to choose a meaningful and feasible level for $s_i$. Therefore, (P1) can be solved to give a reference point for setting $s_i$.

Both (P1) and (P2) are nonlinear programming problems. One well-known method for solving this type of problem is the sequential quadratic programming (SQP) method, well described in [8], [9] and which will be applied here.

## III. Evaluation Results

The calibration signals were created in a hands-free situation with a six-sensor microphone array. The measurements were performed in a Volvo stationwagon. Data were gathered on a multichannel digital audio tape recorder with a sample rate of 12 KHz and a 300–3400-Hz bandwidth. White-noise calibration signals had been used which are emitted individually from the artificial talker and the hands-free loudspeaker as the source and the interference calibration signals, respectively. Interference signals were recorded by emitting an independent sequence of white noise, from the hands-free loudspeaker alone, within the bandwidth. This recording serves as the point source interence calibration signal. Recordings with real speech signals were recorded both individually and while driving. In order to gather background noise signals, the car was driving at a speed of 110 km/h on a paved road. The duration of these signals was 8 sec. More details of the car environment and how the experiment was conducted can be found in [3].

Using these calibration signals, the new method is compared with the LS and the SNIR method described in [3] by designing finite-impluse response (FIR) filters with different lengths.

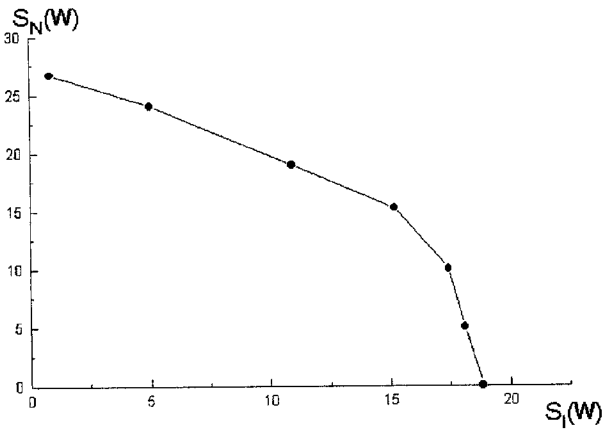


Fig. 1. Different noise and interference suppression levels.

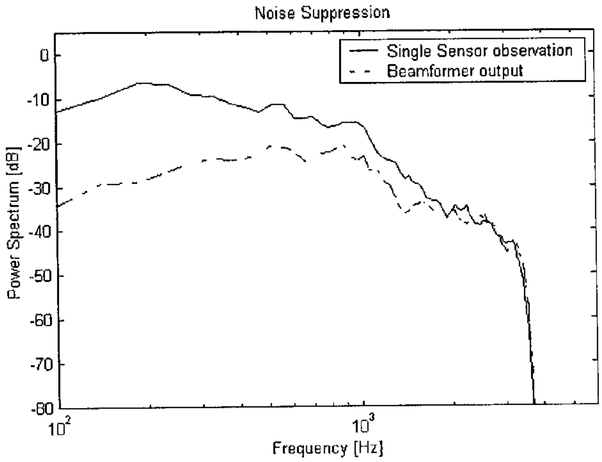


Fig. 2. Spectral estimates using Welch's periodogram of an unprocessed single-microphone observation and the beamformer output signal, when the noise source is active alone.

In Tables I and II, the distortion level $d$ for the new method was fixed at $-25$ dB. (P1) was solved to give an indication to the total suppression level. Then, (P2) was solved with $s_i$ adjusted to give roughly equal suppression levels for both noise and interference. Different initial weights have been tried, and they all yielded similar results. From the tables, LS is poor in the echo suppression, which improves slowly with the filter length, while SNIR is particularly poor in the high distortion level, which stays the same even with a longer filter.

In order to understand the bicriteria objective in the noise and interference suppression, the Pareto optimum set was constructed by solving (P2) with varying $s_i$, where 16-tap FIR filters were used and the distortion level was fixed at $-25$ dB. From Fig. 1, it is observed that the maximum total suppressions occurs at around the level where the noise suppression and the interference suppression are similar. The power spectrum of the optimal beamformer with 16-tap FIR filters and similar suppression levels is shown in Figs. 2 and 3. Very good suppressions for all frequencies are achieved.

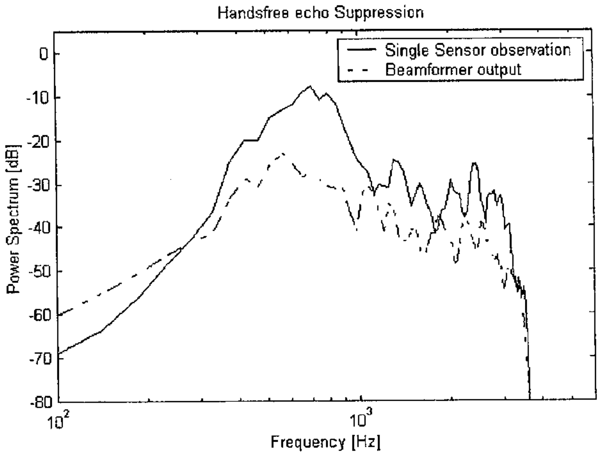


Fig. 3. Spectral estimates using Welch's periodogram of an unprocessed single-microphone observation and the beamformer output signal, when the echo signal is active alone.

## IV. Conclusions

A new design method for broadband microphone arrays has been proposed. The problem has been formulated as a multicriteria optimal beamformer design problem, and the nonlinear programming technique has been applied to tackle it. This method yields improved performance for short FIR filters compared to LS and SNIR methods. The given performance criteria can easily be modified by reformulating the constraints. This will result in a very flexible optimization, in situations where the speech input is used for speech communication as well as speech recognition. We can, thus, have different sets of weights which are optimized according to the best criteria. This will, however, be a topic for further study.

## References


[1] M. Dahl and I. Claesson, "Acoustic noise and echo cancelling with microphone array," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1518–1526, Sept. 1999.

[2] S. Nordholm, I. Claesson, and M. Dahl, "Adaptive microphone array employing calibration signals: An analytical evaluation," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 241–252, May 1999.

[3] S. Nordholm, I. Claesson, and N. Grbić, "Optimal and adaptive microphone arrays for speech input in automobiles," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. Berlin: Springer Verlag, 2001.

[4] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," in *Proc. ASSP-34*, 1986, pp. 1391–1400.

[5] M. M. Sondhi and G. W. Elko, "Adaptive optimization of microphone arrays under a nonlinear constraint," in *Proc. ICASSP*, 1986, pp. 981–984.

[6] M. Zeleny, *Multiple Criteria Decision Making*. New York: McGraw-Hill, 1982.

[7] T. L. Vincent and W. J. Grantham, *Optimality in Parametric Systems*. New York: Wiley, 1981.

[8] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.

[9] K. L. Teo, C. J. Goh, and K. H. Wong, *A Unified Computational Approach to Optimal Control Problems*. White Plains, NY: Longman Scientific and Technical, 1991.