Speech enhancement using an equivalent source inverse filtering-based microphone array

Mingsian R. Bai,[a] Kur-Nan Hur, and Ying-Ting Liu
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

(Received 6 August 2009; revised 16 December 2009; accepted 20 December 2009)

This paper presents a microphone array technique aimed at enhancing speech quality in a reverberant environment. This technique is based on the central idea of single-input-multiple-output equivalent source inverse filtering (SIMO-ESIF). The inverse filters required by the time-domain processing in the technique serve two purposes: de-reverberation and noise reduction. The proposed approach could be useful in telecommunication applications such as automotive hands-free systems, where noise-corrupted speech signal generally needs to be enhanced. SIMO-ESIF can be further enhanced against uncertainties and perturbations by including an adaptive generalized side-lobe canceller. The system is implemented and validated experimentally in a car. As indicated by numerous performance measures, the proposed system proved effective in reducing noise in human speech without significantly compromising the speech quality. In addition, listening tests were conducted to assess the subjective performance of the proposed system, with results processed by using the analysis of variance and a *post hoc* Fisher's least significant difference (LSD) test to assess the pairwise difference between the noise reduction (NR) algorithms. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3291684]

PACS number(s): 43.60.Ac, 43.60.Dh, 43.60.Fg, 43.60.Np [EJS] Pages: 1373–1380

I. INTRODUCTION

Signal processing using microphone arrays has found applications in teleconferencing, telecommunication, speech recognition, speech enhancement, hearing aids, and so forth.[1] In these applications, how to effectively communicate in noisy or reverberant environments has been one of the pressing issues. Array techniques such as the well known delay-and-sum (DAS) beamformer do not function well in such environments due to multiple reflections.[2] To enhance interference rejection, a superdirective beamformer was suggested for its excellent spatial filtering capability.[3] Adaptive arrays provided useful alternatives for interference rejection.[1–13] The generalized side-lobe canceller (GSC) is an elegant approach in beamformer design in which a constrained optimization problem is converted into an unconstrained one from a linear algebraic perspective.[4] Its idea originated from linearly constrained minimum variance (LCMV) beamformer[5] and was first utilized by Owsley.[6] Griffiths and Jim[7] analyzed them and coined the term GSC. Their GSC corresponds to what we called in the present paper the Griffiths–Jim beamformer (GJBF). The key notion of GSC hinges at the use of the blocking matrix that further enhances the performance achievable by the fixed beamformer regardless of how it is implemented. However, oftentimes, adaptive implementation of GSC is preferred for two practical reasons. First, a fixed GSC requires the knowledge of covariance matrix, which is computationally prohibitive in real-time applications. Second, adaptive GSC is more robust in the presence of background noises, pointing errors, and other system uncertainties. There are various forms of adaptive implementation. Among them, GJBF is perhaps the most well known, which eliminates the need for on-line calculation of the signal correlation matrix and is robust against uncertainties and perturbations of the system. However, the GJBF still could fail in a reverberant environment in which multiple reflections cause problems of beamforming.[8] One example of such environment is the interior of a car cabin that is notoriously known to be a noisy and reverberant environment for speech communication. The speech signals tend to be corrupted by noises from the engine, tire, wind, etc., and reflections from the window, dashboard, seats, ceiling, etc.

In this paper, a microphone array technique is proposed for processing speech signals in noisy and reverberant environments. The idea of the technique originated from an equivalent source inverse filtering (ESIF) technique[9] developed for noise source identification purposes. However, the model in this paper is based on a single-input multiple-output (SIMO) structure, while the previous paper is based on a multiple-input multiple-output (MIMO) structure. This seemingly minor difference leads to many distinctive features in the implementation. First, unlike MIMO approach, the SIMO-ESIF formulation results in only one single focus, which simplifies tremendously the filter design that requires only simple phase conjugation and scaling, without having to design explicitly complicated inverse filters. The propagation matrices are basically represented by a vector $\mathbf{h}$, which renders the term $\mathbf{h}^H\mathbf{h}$ in the inverse filter a nonzero scalar and regularization is literarily unnecessary. Second, by posing the problem within a SIMO framework, inverse filters are designed based on measured acoustical plants, or systems to be controlled,[14] that include the effects of not only direct propa-

[a]Author to whom correspondence should be addressed. Electronic mail: msbai@mail.nctu.edu.tw

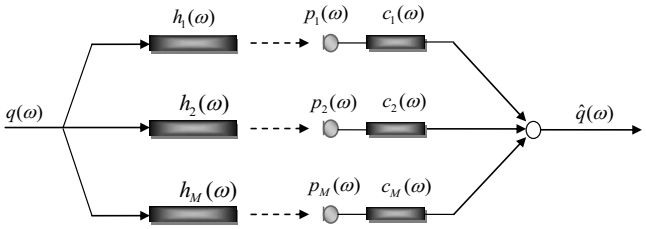


FIG. 1. The block diagram of the SIMO-ESIF algorithm. The parameter $q_m(\omega)$ is the input source, $h_m(\omega)$ is the propagation matrix, and $c_m(\omega)$ is the inverse filters.

gation from the source but also the reflections from the boundaries, which is different from the previous MIMO nearfield equivalence source imaging (NESI) that employs only free-field point source model. It follows that, with inverse filtering, both noise reduction and de-reverberation are fulfilled at the same time. Finally, the present work and the previous approach have totally different purposes in application. The ESIF algorithm aims at speech enhancement for telecommunication, whereas the previous NESI algorithm is intended for noise source identification (NSI) and sound field visualization. Another unique feature of the proposed technique is the use of the GSC to further enhance the performance of the proposed technique. An exact blocking matrix (BM) differing from those used in traditional beamformers is derived in this paper by solving a LCMV problem with two mutually orthogonal subspaces.[5] The leaky least-mean-squares (LMS) algorithms is exploited for adaptive filtering in the multiple-input canceller (MC).[15,16]

The proposed algorithms were implemented for enhancing speech communication quality in a car by using a multichannel data acquisition system. Objective tests were carried out to evaluate the algorithms.[1] In addition, listening tests were conducted to assess the subjective performance of the algorithms with data processed by the multivariate analysis of variance (MANOVA) (Ref. 17) and the least significant difference (Fisher's LSD) *post hoc* test.

II. EQUIVALENT SOURCE INVERSE FILTERING

The central idea of the proposed SIMO-ESIF algorithm is introduced in this section. In Fig. 1, M microphones are employed to pick up the sound emitting from a source positioned in the farfield. In the frequency domain, the sound pressure received at the microphones and the source signal can be related by a $M \times 1$ transfer matrix $\mathbf{H}$ as follows:

$$\mathbf{P} = \mathbf{H}q(\omega), \tag{1}$$

where $q(\omega)$ is the Fourier transform of a scalar source strength, $\mathbf{P}=[p_1(\omega)\cdots p_M(\omega)]^T$ is the pressure vector with T denoting matrix transpose, and $\mathbf{H}=[h_1(\omega)\cdots h_M(\omega)]^T$ is the $M \times 1$ propagation matrix. The aim here is to estimate the source signal $q(\omega)$ based on the pressure measurement $\mathbf{P}$ by using a set of inverse filters

$$\mathbf{C} = [c_1(\omega)\cdots c_M(\omega)]^T, \tag{2}$$

such that $\mathbf{C}^T\mathbf{H} \approx \mathbf{I}$ and therefore

$$\hat{q} = \mathbf{C}^T\mathbf{P} = \mathbf{C}^T\mathbf{H}q \approx q. \tag{3}$$

On the other hand, this problem can also be written in the context of the following least-squares optimization problem:

$$\min_q \|\mathbf{P} - \mathbf{H}q\|_2^2, \tag{4}$$

where $\|\ \|_2$ denotes vector 2-norm. This is an overdetermined problem whose least-squares solution is given by

$$\hat{q} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{P} = \frac{\mathbf{H}^H\mathbf{P}}{\|\mathbf{H}\|_2^2}, \tag{5}$$

where the superscript H denotes Hermitian transpose. Comparison of Eqs. (3) and (5) yields the following optimal inverse filters:

$$\mathbf{C}^T = \frac{\mathbf{H}^H}{\|\mathbf{H}\|_2^2}. \tag{6}$$

If the scalar $\|\mathbf{H}\|_2^2$ is omitted, the inverse filters above reduce to the "phase-conjugated" filters, or the "time-reversed" filters in the free-field context. Specifically, for a point source in the free field, it is straightforward to show that

$$\|\mathbf{H}\|_2^2 = \sum_{m=1}^{M} \frac{1}{r_m^2}, \tag{7}$$

where r_m is the distance between source and the mth microphone. Since $\|\mathbf{H}\|_2^2$ is a frequency-independent constant, the inverse filters and the time-reversed filters differ only by a constant. In a reverberant environment, these filters are different in general. Being able to incorporate the reverberant characteristics in the measured acoustical plant model represents an advantage of the proposed approach over conventional methods such as the DAS beamformer.

In real-time implementation, the inverse filters are converted to the time-domain finite-impulse-response (FIR) filters with the aid of inverse fast Fourier transform (IFFT) and circular shift. Thus, the source signal can be recovered by filtering the pressure signals with the inverse filters $\mathbf{c}(k)$ as follows:

$$\hat{q}(k) = \mathbf{c}^T(k) * \mathbf{p}(k), \tag{8}$$

where k is discrete-time index, $\mathbf{c}(k)$ is the impulse response of the inverse filter, and " *" denotes convolution.

III. ADAPTIVE GSC-ENHANCED SIMO-ESIF ALGORITHM

The SIMO-ESIF algorithm can be further enhanced by introducing an adaptive GSC to the system. The benefit is twofold. The directivity of the array is increased by suppressing the interferences due to side-lobe leakage. Robustness of the array is improved in the face of uncertainties and perturbations. The block diagram of the GSC with M microphones is shown in Fig. 2. The system comprises a fixed beamformer (FBF), a MC, and a BM.[11] The FBF aims at forming a beam in the look direction so that the target signal is passed and signals at other directions are rejected. The $p_m(k)$ is the signal received at the mth microphones and $\hat{q}(k)$ is the output signal of the FBF at the time instant k. The MC consists of

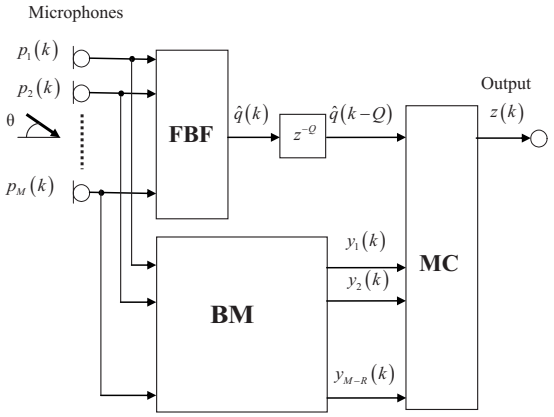


FIG. 2. The block diagram of the GSC, comprised of the FBF, the BM, and the MC.

multiple adaptive filters that generate replicas of components correlated with the interferences. The components correlated with the output signals $y_m(k)$ of the BM are subtracted from the delayed output signal $\hat{q}(k-Q)$ of the FBF, where Q is the number of modeling delay. Contrary to the FBF that produces a main-lobe, the BM forms a null in the look direction so that the target signal is suppressed and all other signals are passed though, hence the name "blocking matrix." The GSC subtracts the interferences that "leak" to the side-lobes in the subtractor output $z(k)$ and effectively improves spatial filtering.

A. Formulation of the blocking matrix

The purpose of the GSC depicted in Fig. 3 lies in minimizing the array output power, while maintaining the unity gain at the look direction (0-deg broadside is assumed here), which can be posed in the following constrained optimization formalism:[5]

$$\min_{\mathbf{w}} E\{|z|^2\} = \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{pp} \mathbf{w} \tag{9}$$

subject to

$$\mathbf{H}^H \mathbf{w} = 1, \tag{10}$$

where z is the array output signal, $\mathbf{R}_{pp}=E\{\mathbf{p}\mathbf{p}^H\}$ is the data correlation matrix, $E\{\}$ symbolizes the expected value, $\mathbf{H}$ is the frequency response vector corresponding to the propagation paths from the source to each microphone, and $\mathbf{w}$ is coefficient vector of the array filters. This constrained optimization problem can be converted into an unconstrained one by decomposing the optimal filter $\mathbf{w}$ into two linearly independent components belonging to two mutually orthogonal subspaces: the constraint range space $R(\mathbf{H})$ and the orthogonal null space $N(\mathbf{H}^H)$.

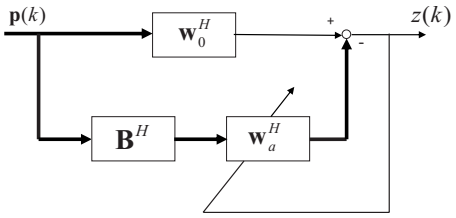


FIG. 3. The block diagram of the SIMO-ESIF-GSC algorithm. The parameter $\mathbf{p}(k)$ is the microphone signal, $\mathbf{B}^H$ is the BM, and $\mathbf{w}_0^H$ is the FBF.

$$\mathbf{w} = \mathbf{w}_0 - \mathbf{v}, \tag{11}$$

where $\mathbf{w}_0 \in R(\mathbf{H})$ is a fixed filter and $\mathbf{v}=\mathbf{B}\mathbf{w}_a \in N(\mathbf{H}^H)$ with $\mathbf{w}_a$ being an adaptive filter. It follows that

$$\mathbf{H}^H \mathbf{w} = \mathbf{H}^H(\mathbf{w} - \mathbf{B}\mathbf{w}_a) = \mathbf{H}^H \mathbf{w}_0 - \mathbf{H}^H \mathbf{B}\mathbf{w}_a \approx 1. \tag{12}$$

The fixed filter $\mathbf{w}_0$ represents the quiescent component that guarantees the essential performance of beamforming. The filter design is off-line since it is independent of the data correlation matrix. It turns out that the minimization can then be carried out in the orthogonal subspace ($\mathbf{v}$) without impacting the constraint.

Traditionally, various *ad hoc* blocking matrices have been suggested. These matrices are based on the simple idea that, for free-field plane waves incident from the farfield broadside direction, $\mathbf{H}=[1\ \ 1\cdots 1]^H$. Since $\mathbf{H}^H\mathbf{B}=\mathbf{0}$, blocking is ensured if the columns of $\mathbf{B}$ sum up to zero; e.g., subtraction of signals of adjacent channels is a widely used approach. However, for a complex propagation matrix in a reverberant field, these *ad hoc* blocking matrices are inadequate. As a major distinction between the present approach and the conventional approaches, we shall derive an exact blocking matrix for a more general acoustical environment.

To fulfill the condition that $\mathbf{B}\mathbf{w}_a \in N(\mathbf{H}^H) \Leftrightarrow \mathbf{H}^H\mathbf{B}\mathbf{w}_a=0$, the columns of $\mathbf{B}$ must be constructed from the basis vectors of $N(\mathbf{H}^H)$ such that $\mathbf{H}^H\mathbf{B}=\mathbf{0}$.

$$\text{Let } \mathbf{H} = [a_1, a_2, \ldots, a_n]^H, \quad \mathbf{x} = [x_1, x_2, \ldots, x_n] \in N(\mathbf{H}^H)$$

$$\mathbf{H}^H\mathbf{x} = 0 \Rightarrow a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$$

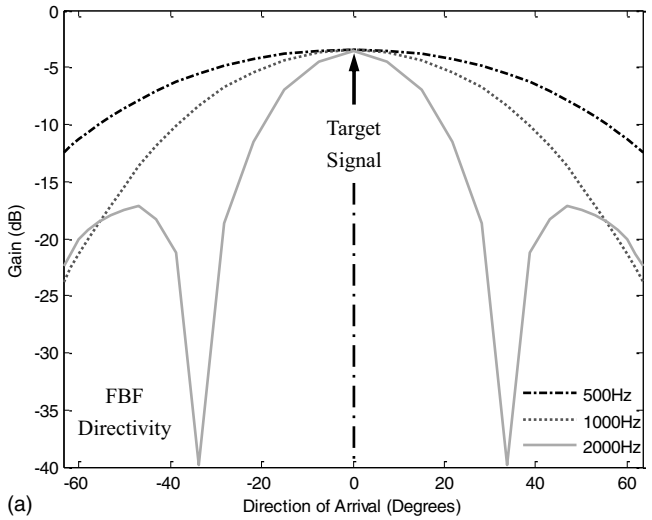
$$\text{If } a_1 \neq 0, \quad x_1 = -\frac{a_2}{a_1}x_2 - \frac{a_3}{a_1}x_3 - \cdots - \frac{a_n}{a_1}x_n$$

$$\text{Let } x_2 = \alpha_2, \quad x_3 = \alpha_3, \ldots, x_n = \alpha_n$$

$$\Rightarrow x_1 = -\frac{a_2}{a_1}\alpha_2 - \frac{a_3}{a_1}\alpha_3 - \cdots - \frac{a_n}{a_1}\alpha_n$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \alpha_2 \underbrace{\begin{bmatrix} -\frac{a_2}{a_1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{v}_2} + \alpha_3 \underbrace{\begin{bmatrix} -\frac{a_3}{a_1} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{v}_3} + \cdots + \alpha_n \underbrace{\begin{bmatrix} -\frac{a_n}{a_1} \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_n}.$$

It is not difficult to see that $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ are linearly independent and form the basis of the null space $N(\mathbf{H}^H)$. Thus, the matrix $\mathbf{B}=(\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n)$ comprised of $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ as its columns can be used as the blocking matrix; i.e.,



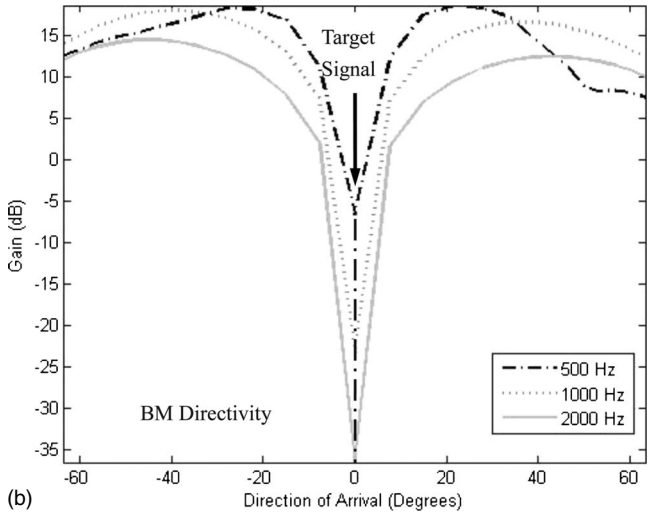


FIG. 4. The directivity pattern of the SIMO-ESIF-GSC algorithm at difference frequencies. (a) FBF with a main-lobe at the look direction. (b) BM with a null at the look direction.

$$\mathbf{B} = \begin{bmatrix} -\frac{a_2}{a_1} & -\frac{a_3}{a_1} & \cdots & -\frac{a_n}{a_1} \\ 1 & 0 & & 0 \\ 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{13}$$

Physical insights can be gained by observing the beam patterns of the FBF and the BM shown in Fig. 4. Three sine wave signals at 500 Hz, 1 kHz, and 2 kHz are used to compare the performance of the BM between FBF, respectively. In the look direction, the FBF forms a main-lobe, whereas the BM forms a null so that the signal in the look direction is "blocked." The blocked path will attempt to further reduce the noise or interference outside the principal look direction (side-lobes). Note that the formulation above is in the frequency domain. For real-time implementation, the blocking matrix **B** needs to be converted to impulse responses using IFFT and circular shift.

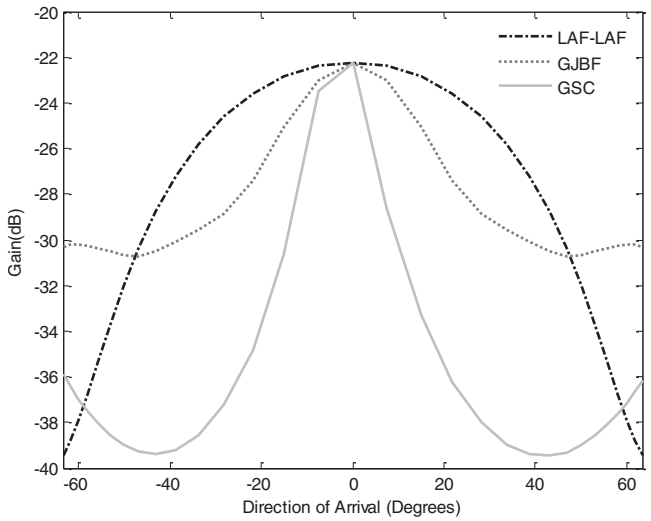


FIG. 5. The comparison of the beam patterns at 500 Hz obtained using the GJBF, LAF-LAF, and SIMO-ESIF-GSC algorithms.

B. Multiple-input canceller

In practice, the GSC is implemented using adaptive filters that are generally more robust than fixed filters. The need to compute the data correlation matrix $\mathbf{R}_{pp}$ is eliminated using such approach. For example, the leaky adaptive filters (LAFs) (Ref. 16) can be used in the MC block. LAFs subtract the components correlated with $y_n(k)$, $(n=0,\ldots,N)$ from $\hat{q}(k-Q)$, where Q is the modeling delay for causality. Let M_2 be the number of taps in each LAF and $\mathbf{w}_n(k)$ and $\mathbf{y}_n(k)$ be the coefficient vector and the signal vector of the nth LAF, respectively. The output of the MC module can be written as

$$z(k) = \hat{q}(k-Q) - \sum_{n=0}^{N-1} \mathbf{w}_n^T(k)\mathbf{y}_n(k), \tag{14}$$

$$\mathbf{w}_n(k) \triangleq [w_{n,0}(k), w_{n,l}(k), \ldots, w_{n,M_2-1}(k)]^T, \tag{15}$$

$$\mathbf{y}_n(k) \triangleq [y_n(k), y_n(k-1), \ldots, y_n(k-M_2+1)]^T. \tag{16}$$

The filter coefficients can be updated using the LMS algorithm

$$\mathbf{w}_n(k+1) = \mathbf{w}_n(k) + \mu z(k)\mathbf{y}_n(k), \tag{17}$$

where μ is the step size.

In Fig. 5, the beam pattern at 500 Hz of the proposed adaptive GSC algorithm is compared with two other conventional algorithms: GJB (Ref. 7) and LAF-LAF.[16] The GJBF algorithm adopts subtracted signals of adjacent channels as its BM block, whereas LAF-LAF algorithm uses adaptive filters to block the target signals. Both algorithms use adaptive algorithm identical to the MC block in Eq. (17). Clearly seen in Fig. 5, the proposed adaptive GSC algorithm attains the sharpest beam in the look direction with minimum side-lobes.

IV. ARRAY PERFORMANCE MEASURES

In the section, objective performance measures are defined for evaluating the array performance.[1] With the first microphone as the reference, the input signal to noise ratio (SNR) is defined as

$$\mathrm{SNR}_1(\mathrm{dB}) = 10\log\frac{E\{x_1(k)^2\}}{E\{v_1(k)^2\}}, \tag{18}$$

where k is the discrete-time index, and $x_1(k)$ and $v_1(k)$ are the speech signal and the noise, respectively, received at microphone 1. The output SNR can also be defined for the array output

$$\mathrm{SNR}_\mathrm{A}(\mathrm{dB}) = 10\log\frac{E\{|\mathbf{c}(k)^T * \mathbf{x}(k)^2|\}}{E\{|\mathbf{c}(k)^T * \mathbf{v}(k)^2|\}}, \tag{19}$$

where $\mathbf{c}(k)$ is the impulse response of the inverse filter and " *" denotes convolution. Hence, the SNR gain is obtained by subtracting the output SNR from the input SNR.

$$\mathrm{SNR}(\mathrm{dB}) = \mathrm{SNR}_\mathrm{A} - \mathrm{SNR}_1. \tag{20}$$

The SNRG quantifies the noise reduction performance due to array processing. However, noise reduction comes at the price of speech distortion in general. To assess speech distortion, a speech-distortion index (SDI) is defined as

$$\mathrm{SDI}(\mathrm{dB}) = 10\log\frac{E\{x_1(k)^2\}}{E\{|x_1(k) - \mathbf{c}(k)^T * \mathbf{x}(k)|^2\}}. \tag{21}$$

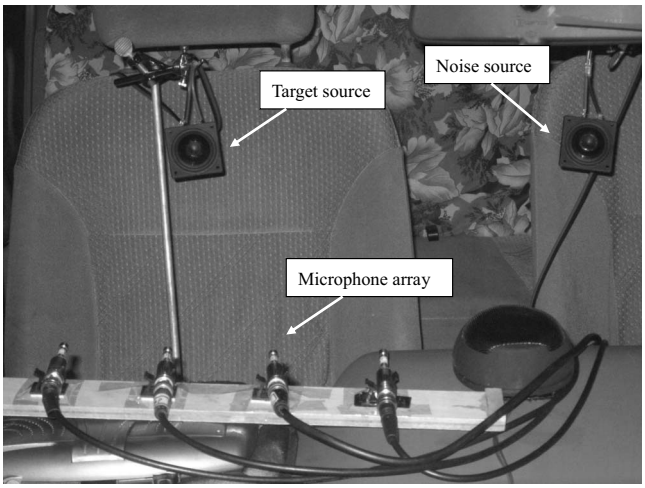
It is impractical to maximize both indices at the same time. The aim of array processing is then to reach the best compromise between the two indices.

V. OBJECTIVE AND SUBJECTIVE PERFORMANCE EVALUATIONS

The proposed algorithms have been examined experimentally in the vehicle compartment of a 2-l sedan. Figure 6 shows the experimental arrangement inside the car. Array signal processing algorithms are all implemented on National Instruments LABVIEW 8.6 and NI-PXI 8105 data acquisition system.[18] The sampling rate is 8 kHz. The sound pressure data were picked up by using a four-microphone (PCB 130D20) linear uniform array with inter-element spacing of 0.08 m. A loudspeaker positioned at (0.4 m, 0-deg) with respect to the array center was used to broadcast a clip of male speech in English, while another loudspeaker positioned at (0.3 m, 53-deg) was used to generate white noise as the interference.

Objective and subjective experiments were undertaken to evaluate the proposed methods. The SIMO-ESIF algorithm is used as the FBF and 512-tapped adaptive filters with step size μ=0.001 are used in the MC and LAF. There are variations of the SIMO-ESIF algorithm, depending on the plant model used and the filtering method in the FBF, as summarized in Table I. Two kinds of plant models, the free-field point source model and the measured plant in the car, are employed for designing the inverse filters. Two filtering methods, the inverse filtering and the time-reversed filtering, are employed in the FBF design. In addition, three variations of the processing methods with GSC are also included in Table I.

(a)



(b)

FIG. 6. The experimental arrangement for validating the SIMO-ESIF algorithms. (a) The test car. (b) The experimental arrangement inside the car.

A. Objective evaluation

The objective measures SNR_1, SNR_A, SNRG, and SDI are employed to assess the performance of six proposed algorithms. The experimental results are summarized in Table II. By comparing the SIMO-ESIF and the SIMO-ESIF-GSC algorithms, the algorithms with GSC have attained signifi-

TABLE I. The acronyms and descriptions of six SIMO-ESIF algorithms.

Algorithm	Acronym	Description
SIMO-ESIF	PIF	Point source model-based inverse filtering
	MIF	Measured plant-based inverse filtering
	MTR	Measured plant-based time-reversed filtering
	GSC-PIF	Point source model-based inverse filtering
SIMO-ESIF-GSC	GSC-MIF	Measured plant-based inverse filtering
	GSC-MTR	Measured plant-based time-reversed filtering

TABLE II. The objective performance summary of the six algorithms.

	PIF		MIF		MTR	
SIMO-ESIF GSC	Without	With	Without	With	Without	With
SNR_I(dB)	3.79	3.79	3.79	3.79	3.79	3.79
SNR_A(dB)	12.96	15.28	15.56	19.19	13.58	13.66
SNRG(dB)	9.16	11.49	11.77	15.41	9.78	9.87
SDI(dB)	2.87	2.60	1.72	1.59	0.86	1.56

cantly higher noise reduction (SNRG) and lower speech distortion (SDI) than the algorithms without GSC. The time-reversed filters in general yield inferior performance than the inverse filters. The inverse filtering with the measured plant model considerably outperforms the point source model, e.g., SNRG of GSC-MIF=15.41 dB vs SNRG of GSC-PIF =11.49 dB. The implication of this result is that the inverse filters based on measured plant models have provided "dereverberation" effect in addition to noise reduction. Although the point source model-based inverse filtering (PIF) method tends to yield the least distortion, its noise reduction performance is also the worst. Figure 7 compares the time-domain wave forms obtained using SIMO-ESIF algorithm with and without GSC. Evidently, introduction of GSC has positive impact on noise reduction performance of the array.

Table III compares the proposed adaptive GSC algorithm and two other conventional algorithms, GJBF (Ref. 7) and CCAF.[11] The GJBF algorithm subtracts signals of adjacent channels as its BM block, whereas coefficient-constrained adaptive filtering (CCAF) algorithm uses constrained adaptive filters to block the target signals. Both algorithms use the adaptive algorithm identical to the MC block. The result revealed that the SIMO-ESIF algorithm augmented with the GSC outperformed the SIMO-ESIF algorithm without GSC. Among the GSC-based algorithms, the proposed GSC had attained the highest SNRG performance. The proposed GSC algorithm performed the best in noise reduction.

B. Subjective evaluation

Apart from the preceding objective tests, listening tests were conducted according to the ITU-R BS1116 (Ref. 19) to validate the algorithms. Subjective perception of the proposed algorithms was evaluated in terms of noise reduction and speech distortion. Specifically, three subjective attributes including signal distortion (SIG), background intrusiveness (BAK), and overall quality (OVL) were employed in the test. Fourteen participants in the listening tests were instructed with definitions of the subjective attributes and the procedures prior to the test. The subjective attributes are measured on an integer scale from 1 to 5. The participants were asked to respond in a questionnaire after listening. The six proposed algorithms previously used in the objective test are compared in the listening test. The test signals and conditions remain the same as in the preceding objective tests. A reference signal and an anchor signal are required in the ITU-R BS1116. In the test, the unprocessed signal received at the first microphone was used as the reference, while the lowpass-filtered reference was used as the hidden anchor. The mean and spread of the results of listening test are illustrated in Fig. 8. In order to access statistical significance of the results, the test data were processed using MANOVA (Ref. 17) with significance levels summarized in Table IV. Cases with significance levels below 0.05 indicate that statistically significant difference exists among the algorithms. In particular, the difference of the indices SIG and BAK

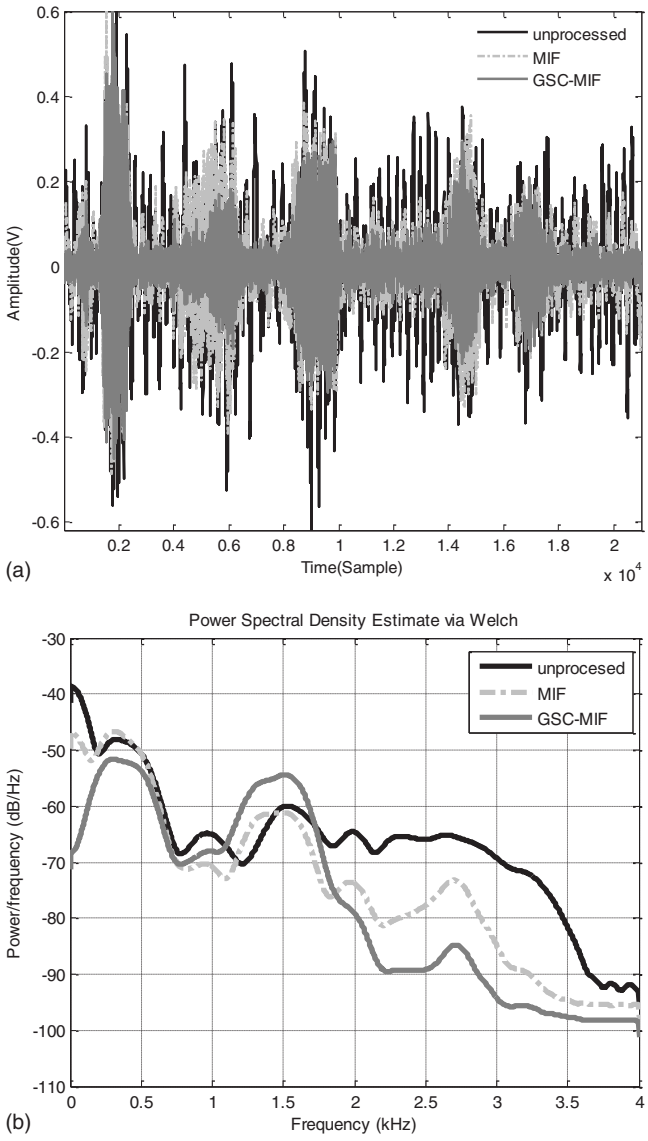


FIG. 7. The comparison of SIMO-ESIF-MIF algorithm and SIMO-ESIF-GSC-MIF algorithm by experimental measurement. (a) The time-domain wave forms. (b) The power spectral density functions.

TABLE III. The objective performance summary of the four beamforming algorithms including the ESIF, ESIF-GSC, GJBF, and CCAF algorithms.

	MIF			
Objective index	ESIF	ESIF-GSC	GJBF	CCAF
SNR_1(dB)	−1.04	−1.04	−1.04	−1.04
SNR_A(dB)	6.20	12.72	10.27	9.92
SNRG(dB)	7.24	13.76	11.31	10.96
SDI(dB)	1.86	1.42	2.49	1.90

among the six proposed methods was found to be statistically significant. Multiple regression analysis was applied to analyze the linear dependence of the OVL on the SIG and BAK. The regression model was found to be OVL=1.71+0.2 ×SIG+0.28×BAK. It revealed that the SIG has comparable but only slightly higher influence on the OVL than the BAK, whereas the indices SIG and the BAK are normally trade-offs. This explains why no significant difference can be found among methods in the OVL.

After the MANOVA, a *post hoc* Fisher's LSD test was employed to perform multiple paired comparisons. In Fig. 8, as opposed to the results of objective evaluation, the GSC-MIF algorithm performed not quite as expected in SIG. The price paid for the high noise reduction seems to be the signal distortion, which was noticed by many subjects. For the SIG index, the results of the *post hoc* test reveal that the GSC-PIF method outperforms the other methods. For the BAK index, the GSC-MIF method received the highest grade among all methods, which means that the inverse filtering approach has achieved both de-reverberation and noise reduction successfully. Despite the excellent performance in SIG, the PIF algorithm received low scores in BAK, which is consistent with the observation in the objective test. On the other hand, the GSC-PIF algorithm received higher SIG grade than plain PIF algorithm, indicating the GSC algorithm enhanced the SIMO-ESIF algorithm. However, the grades in the SIG and BAK indices showed no significant difference between the measured plant-based time-reversed filtering (MTR) and

TABLE IV. The MANOVA output of the listening test of the six algorithms. Cases with significance value p below 0.05 indicate that statistically significant difference exists among all methods.

	Significance value p		
Noise type	SIG	BAK	OVL
White noise	0.000	0.000	0.847

GSC-MTR algorithms. By comparing the BAK grade, all the proposed methods performed better than the reference signal.

Figure 9 compares the proposed GSC algorithm, GJBF,[6] and CCAF (Ref. 11) algorithms. The proposed GSC algorithm attained the highest BAK grades, while it also yielded lower SIG grades than the other algorithms. Apparently, the proposed GSC had attained the best performance in noise reduction at the expense of signal distortion. This is a typical trade-off for speech enhancement algorithms in general one has to face between signal distortion and noise reduction performance.

VI. CONCLUSIONS

A SIMO-ESIF microphone array technique has been developed for noisy automotive environments. Speech communication quality has been improved owing to the noise reduction and de-reverberation functions provided by the proposed system. With the use of specially derived BM of the GSC, the performance of SIMO-ESIF has been further enhanced.

The proposed algorithms have been validated via extensive objective and subjective tests. Overall, the results reveal that both de-reverberation and noise reduction can be achieved by using the SIMO-ESIF techniques. The methods exhibit different degrees in trading off noise reduction performance and speech-distortion quality. The MIF and GSC-MIF algorithms seem to have achieved a satisfactory compromise between these two attributes. All this leads to the conclusion that SIMO-ESIF-GSC-MIF proves effective in reducing noise and interference without markedly compromising speech quality.

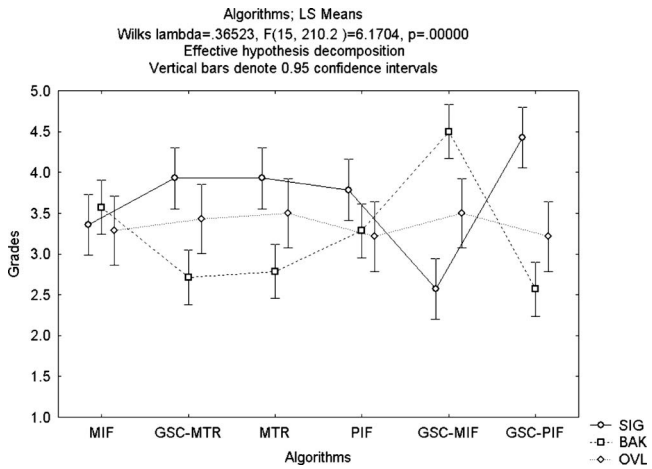


FIG. 8. The MANOVA output of the subjective listening test for the six SIMO-ESIF algorithms. Three subjective attributes including signal distortion (SIG), background intrusiveness (BAK), and overall quality (OVL) were evaluated in the test.

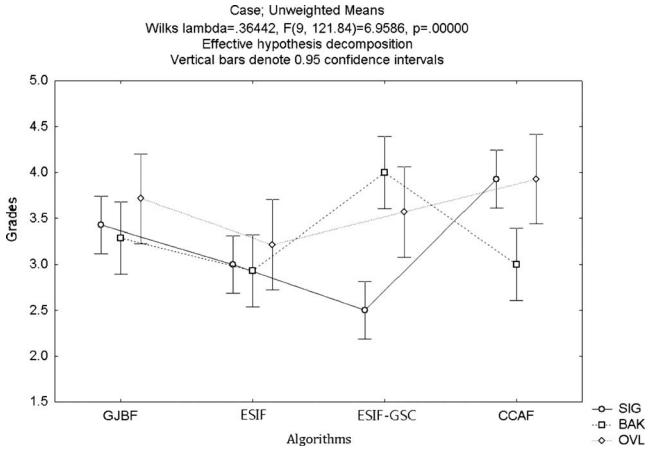


FIG. 9. The MANOVA output of the subjective listening test for the different GSC algorithms. Three subjective attributes including signal distortion (SIG), background intrusiveness (BAK), and overall quality (OVL) were evaluated in the test.

ACKNOWLEDGMENT

The work was supported by the National Science Council of Taiwan, Republic of China, under Project No. NSC 97-2221-E-009-010-MY3.

[1]J. Benesty, J. Chen, and Y. Huang, *Microphone Arrays Signal Processing* (Springer, New York, 2008).
[2]J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multi-microphone noise reduction techniques for hands-free speech recognition—A comparative study," in Robust Methods for Speech Recognition in Adverse Conditions (ROBUST99), Tampere, Finland (1999), pp. 171–174.
[3]J. Bitzer, K. D. Kammeyer, and K. U. Simmer, "An alternative implementation of the superdirective beamformer," Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October (1999).
[4]H. L. Van Trees, *Optimum Array Processing* (Wiley, New York, 2002).
[5]H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1365–1376 (1987).
[6]N. L. Owsley, "Source location with an adaptive antenna array," Technical Report No. AD0719896, Naval Underwater Systems Center, National Technical Information Service, Springfield, VA, 1971.
[7]L. J. Griffiths and C. W. Jim, "An alternative approach to linear constrained adaptive beamforming," IEEE Trans. Antennas Propag. **30**, 27–34 (1982).
[8]J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multichannel noise reduction—Algorithms and theoretical limits," in Proceedings of the EURASIP European Signal Proceeding Conference (EUSIPCO), Rhodes, Greece, September (1998), Vol. **1**, pp. 105–108.
[9]M. R. Bai and J. H. Lin, "Source identification system based on the time-domain nearfield equivalence source imaging: Fundamental theory and implementation," J. Sound Vib. **307**, 202–225 (2007).
[10]M. Brandstein and D. Ward, *Microphone Arrays* (Springer, New York, 2001).
[11]O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone array with a blocking matrix using constrained adaptive filters," IEEE Trans. Signal Process. **47**, 2677–2684 (1999).
[12]Y. Grenier, "A microphone array for car environment," Speech Commun. **12**, 25–39 (1993).
[13]O. L. Frost III, "An algorithm for linearly-constrained adaptive array processing," Proc. IEEE **60**, 926–935 (1972).
[14]G. F. Franklin, M. L. Workman, and D. Powell, *Feedback Control of Dynamic Systems*, 2nd ed. (Addison-Wesley, Boston, MA, 1993).
[15]I. Claesson and S. Nordholm, "A spatial filtering approach to robust adaptive beamforming," IEEE Trans. Antennas Propag. **40**, 1093–1096 (1992).
[16]O. Hoshuyama and A. Sugiyama, "A robust generalized sidelobe canceller with a blocking matrix using leaky adaptive filters," Electron Commun. Jpn. **80**, 56–65 (1997).
[17]S. Sharma, *Applied Multivariate Techniques* (Wiley, New York, 1996).
[18]National Instruments, http://sine.ni.com/nips/cds/view/p/lang/zht/nid/202630 (Last viewed 7/17/2009).
[19]ITU-R Rec. BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunications Union, Geneva, Switzerland, 1994.