01 Jan 2005

A Microphone Array System for Multimedia Applications with Near-Field Signal Targets

Y. Rosa Zheng
Missouri University of Science and Technology, zhengyr@mst.edu

M. El-Tanany

R. A. Goubran

Hongchi Shi

A Microphone Array System for Multimedia Applications With Near-Field Signal Targets

Yahong Rosa Zheng, *Member, IEEE*, Rafik A. Goubran, *Member, IEEE*, Mohamed El-Tanany, and Hongchi Shi, *Senior Member, IEEE*

Abstract—A microphone array beamforming system is proposed for multimedia communication applications using four sets of small planar arrays mounted on a computer monitor. A new virtual array approach is employed such that the original signals received by the array elements are weighted and delayed to synthesize a large, nonuniformly spaced, harmonically nested virtual array covering the frequency band [50, 7000] Hz of the wideband telephony. Subband multirate processing and near-field beamforming techniques are then used jointly by the nested virtual array to improve the performances in reverberant environments. A new beamforming algorithm is also proposed using a broadband near-field spherically isotropic noise model for array optimization. The near-field noise model assumes a large number of broadband random noises uniformly distributed over a sphere with a finite radius in contrast to the conventional far-field isotropic noise model which has an infinite radius. The radius of the noise model, thus, adds a design parameter in addition to its power for tradeoffs between performance and robustness. It is shown that the near-field beamformers designed by the new algorithm can achieve more than 8-dB reverberation suppression while maintaining sufficient robustness against background noises and signal location errors. Computer simulations and real room experiments also show that the proposed array beamforming system reduces beampattern variations for broadband signals, obtains strong noise and reverberation suppression, and improves the sound quality for near-field targets.

Index Terms—Array signal processing, microphone, modeling, multimedia communication, speech enhancement.

I. INTRODUCTION

MULTIPLE microphones and microphone arrays are widely used for their enhanced performances in signal detection and identification [1], [2], source localization [3], [4], noise and interference suppression [5], [6], and sensor networking and multisensor fusion [7], [8]. Microphone array processing falls into two basic categories: One is array beamforming and another is source localization and tracking. Array beamforming utilizes the multiple sensor's input signals to form a directional beam at the desired target, thus, effectively suppresses background noises and interference. It provides a convenient, low-cost means of hands-free sound acquisition with enhanced quality of the desired signal [9]–[15]. Source localization and tracking techniques estimate the signal locations and their movements, which can guide microphone array beamformers, video cameras, and other sensory systems to pick up the desired signals. These two types of array processing techniques may be accomplished by a single microphone array and cooperated with each other in a multimedia communication system.

This paper focuses on microphone array beamforming for speech enhancement and reverberant noise suppression. For multimedia communication applications in small rooms and automobiles, the signal sources are located close to the sensor array and the wave front curvature can be significant within the array's aperture [16]. Conventional far-field beamforming methods may not be suitable because they use the simplified far-field propagation model assuming that all impinging signals at the sensor array are plane waves. As a rule of thumb, when signals are located within the radial distance of R_a^2/λ, where R_a is the size of the array and λ is the wavelength of the operating frequency, far field beamforming can result in severe performance loss and near-field beamforming techniques have to be used [17]–[20].

On the other hand, the microphone array application also imposes constraints on the physical size of the array due to its installation and mounting. The limited size of the array means limited aperture and performance, especially at low frequencies of audio signals. Several approaches have been proposed in the literature to overcome the size limitation and improve the performance. One approach is the super-gain method [12], [13] which trades off the array's robustness against noises and errors with the array directivity at low frequencies through constrained optimization. Another approach has been proposed for hearing aid applications using a microelectromechanical system (MEMS)-based microarrays which electronically add propagation delays to every element of the microarray, thus increasing the array's aperture across the audio signal band [14]. The basic idea is that the desired signal location is known to the array; therefore, propagation delays can be adjusted according to the signal location and operating frequency such that the desired signal received at each array element is first shifted in phase then added constructively to achieve the highest gain at the array output. Signals from other locations are attenuated due to destructive combining by the array beamformer. To cover the wide acoustic frequency band, uniformly sampled frequency points have been

used in [14] so that the propagation delay corresponding to half wavelength spacing of the operating frequency is added and constant beam width beamformer is achieved. Far-field beamforming method has also been used in [14] primarily due to the small physical size of the microarray.

In this paper, we combine the optimization design method with the propagation delay method via a new virtual array approach. We propose a physical array system consisting of four planar array sets mounted on the corners of a computer monitor. A nested virtual array is first synthesized by the received signals of the four physical array sets through added propagation delays. Virtual elements of the virtual array are used to illustrate the synthesized propagation delay effect as if the elements are spaced far apart at the half (or smaller) wavelength of the operating frequency. Subband near-field beamformers are then designed by a new optimization method based on the synthesized virtual array. The proposed virtual array approach has two advantages. First, the locations of the virtual elements can help to determine the propagation delays corresponding to the desired signal. This is especially important when near-field signal sources are considered. Second, after the synthesis of the virtual array, many existing beamforming methods may be applied directly to improve the performances.

Different from the traditional uniform sampling methods, this paper uses the harmonically nested virtual array which synthesizes fewer signals at nonuniformly spaced frequency intervals. It then subbands the signals and uses a tapped-delay line for beamforming in each subband. Physically nested arrays have been widely employed for broadband beamforming [5], [6], [9], [10], [27] in microphone array applications. For broadband signals whose high-frequency-to-low-frequency ratio is much larger than 10:1, this approach provides a compromise solution to improving the performance and reducing the system complexity. In our applications, the use of nested virtual array and subband beamforming is similar to the nested physical array approach. The additional advantage is the reduced synthesis errors.

A new near-field beamformer design is also proposed which optimizes the beamformer weights of the virtual arrays using a new broadband, near-field, spherically isotropic noise model [21]. Constrained optimization for near-field beamformer design is more robust against location errors than dynamically adaptive beamforming methods [11]. It also achieves better noise and reverberation suppression than fixed-weight delay-filter-and-sum beamformers [19], [20]. Conventionally, the far-field spherically isotropic noise model has been used [11], [13], [20] for near-field microphone array optimization, where the noise is assumed to be uniformly distributed on a sphere with infinite radius. This model is more convenient than the image model [26] for simulating reverberant noises because it is independent of the environment and array settings. Recently, a narrow-band, near-field, spherically isotropic noise model [17] has been used in optimization of the near-field beamformer that shows improved performance over the far-field isotropic noise model only when a very small radius of the noise sphere is chosen. In this paper, we extend the near-field noise model to broadband isotropic noises. We show that the performance of the near-field beamformer can be improved

significantly when the radius of the noise sphere is chosen larger than the distance between the array and the target signal. This is particularly suitable for microphone array applications in reverberant enclosures because the reverberant noises are virtually originated from farther distances due to reflections [26]. Furthermore, the radius of the near-field noise sphere provides a design parameter in addition to the power of the noise field for tradeoffs between de-reverberation gain and array robustness. When the new noise model is applied to the low-frequency subbands, our design method results in super-gain beamformers. By carefully choosing the radius and power of the noise sphere, the robustness of the beamformer against errors can be kept within a satisfactory level.

The proposed microphone array system achieves better performances than far-field beamformers and conventional near-field beamformers in terms of distance discrimination for near-field targets, reduced beampattern variations for broadband signals, and strong suppression of reverberant noises. Computer simulations show that the proposed near-field beamformers can improve distance discrimination for near-field targets, reduce beampattern variations for broadband signals, and achieve strong suppression of reverberant noises. Compared to existing beamformer methods, the near-field beamformer optimized by the new noise model improves reverberation suppression by 8 dB in low-frequency subbands, and 1 to 3 dB in the middle- and high-frequency subbands. This is significant because reverberant noises are more prominent at low frequencies and relatively moderate at middle and high frequencies [9]. Meanwhile, it maintains sufficient robustness against background noises and location/synthesis errors. Real room experiments with a perceptual evaluation of speech quality (PESQ) test also shows that the beamformer improves the speech quality in a reverberant room with the objective mean opinion score (MOS) increased by 1.2 points over the reverberant speech.

II. PROPOSED VIRTUAL ARRAY APPROACH

In multimedia communication applications, microphone arrays are generally installed on computer monitors. In our design, four sets of small-sized uniform planar arrays are mounted on the four corners of a 21-in computer monitor, as illustrated in Fig. 1. The desired target is normally located in front of the monitor within a short distance. Each small planar array set has 5×5 elements with interelement spacing of 2.4 cm, which is the half wavelength of 7200 Hz. The size of each planar array set is then 9.6×9.6 cm. The distance between the centers of two planar array sets is 38.4 cm, which equals the half wavelength of 450 Hz. This configuration is to cover the frequency band of [50, 7000] Hz according to the G.722 standard [22].

To obtain sufficient spatial resolution covering the entire acoustical frequency band, a harmonically nested virtual array is synthesized from the original array sets by adding propagation delays to the received signals. The synthesized array consists of several harmonically nested subarrays as illustrated in Fig. 2. Each subarray is a uniform planar array covering an octave subband: from Sub1 covering $B_1 = [3.6, 7.2]$ kHz to Sub7 covering $B_7 = [0.05, 0.1175]$ kHz. The three high-frequency band subarrays Sub1 to Sub3 each have 5×5 elements with
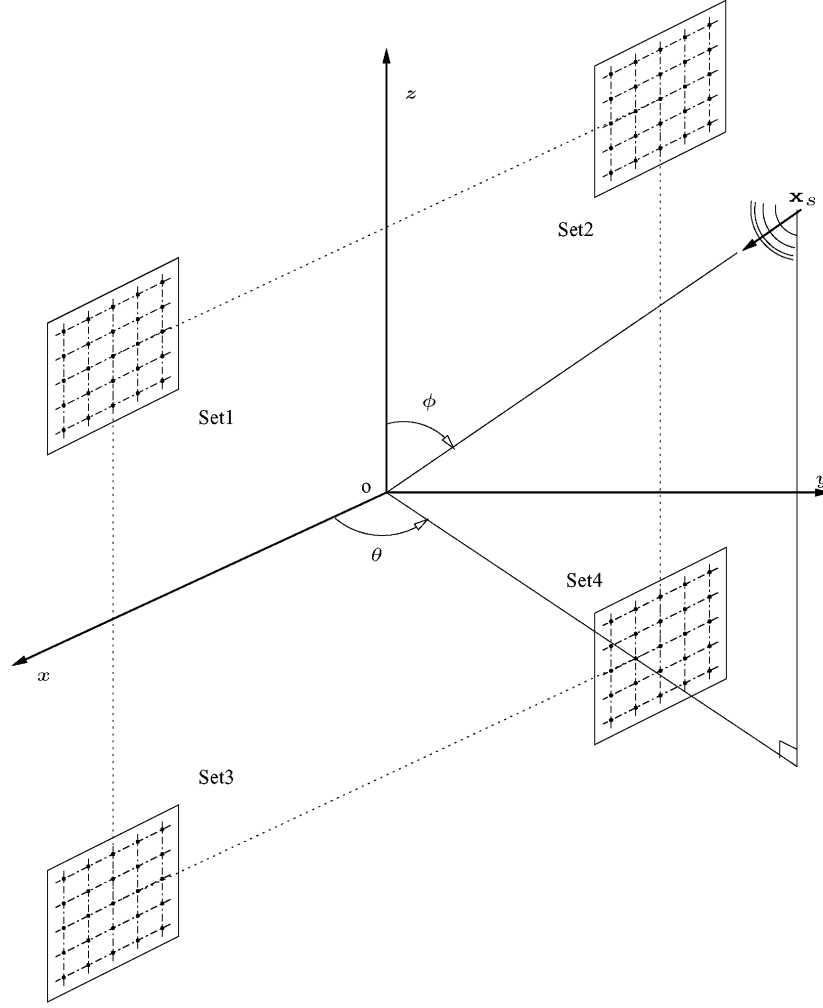
Fig. 1. Four sets of planar arrays mounted on the corners of a computer monitor. Each array set has 5×5 elements with 2.4-cm interelement spacing. The distance between the centers of two array sets is 38.4 cm.

$\lambda/2$ spacing; the middle-frequency band subarrays Sub4 and Sub5 each have 9×9 elements with $\lambda/4$ spacing, where λ is the wavelength of the high-frequency edge of the corresponding subband. The low-frequency band subarrays Sub6 and Sub7 share the elements of Sub5. The reason for quarter-wavelength spacing at Sub4 and Sub5 is that near-field beamforming generally requires smaller spacing to avoid spatial aliasing [18] than the half-wavelength spacing required by far-field beamforming. Besides, it is also found [19] that smaller spacing could result in better performances for near-field beamformers, especially for larger arrays observing greater wave front curvature. The reason for shared elements at Sub5, Sub6, and Sub7 is to reduce the sensitivity to synthesis errors and avoid reduced performance in the very near region of a large array.

The virtual array has a total number of 169 elements synthesized from the 100 original elements. The size of the virtual array is as large as 1.56×1.56 m^2 compared to the original array size of 0.48×0.48 m^2. Consequently, the wave front curvature is significant for signals located within the radial distance of $2D_a^2/\lambda$, where D_a is the size of the subarray [16]. Assume that the signal originates from a known spatial point $\mathbf{x}_s$ within the near field of the virtual array, then the near-field propagation delay and power attenuation have to be considered for synthesis.

The weighted average method is used to synthesize the signals of the virtual array from the ones received by the original array elements. Let the locations of the original array elements be $\mathbf{x}_{i,j}^{(p)}$ and the locations of the virtual array elements be $\hat{\mathbf{x}}_{m,n}$, where the superscript $p = 1, 2, 3, 4$ indicates the planar array set at each corner of the monitor, the subscript pairs (i, j) and (m, n) are the indices of the array elements. The received signals at the original array are denoted as $S_{i,j}^{(p)}(t)$. For subarrays Sub1 to Sub3, the synthesized signal at the (i, j)th element is generated by

$$\hat{S}_{i,j}(t) = \frac{1}{4} \sum_{p=1}^{4} b_{i,j}^{(p)} S_{i,j}^{(p)} \left(t - \frac{\hat{r}_{i,j} - r_{i,j}^{(p)}}{c} \right) \qquad (1)$$

where $i = -2, -1, \cdots, 2$, and $j = -2, -1, \cdots, 2$. The distances are defined as $\hat{r}_{i,j} = |\mathbf{x}_s - \mathbf{x}_{i,j}^{(p)}|$ and $r_{i,j}^{(p)} = |\mathbf{x}_s - \mathbf{x}_{i,j}^{(p)}|$. The propagation speed of the sound wave is $c = 343$ m/s, and $b_{i,j}^{(p)}$ are the weights associated with the distances $\hat{r}_{i,j}$ and $r_{i,j}^{(p)}$ by $b_{i,j}^{(p)} = r_{i,j}^{(p)}/\hat{r}_{i,j}$.

Subarrays Sub4 and Sub5 each have 9×9 elements to be synthesized. Denote the synthesized signal by $\hat{S}_{m,n}(t)$ where $m = -4, -3, \cdots, 4$, and $n = -4, -3, \cdots, 4$. The virtual array
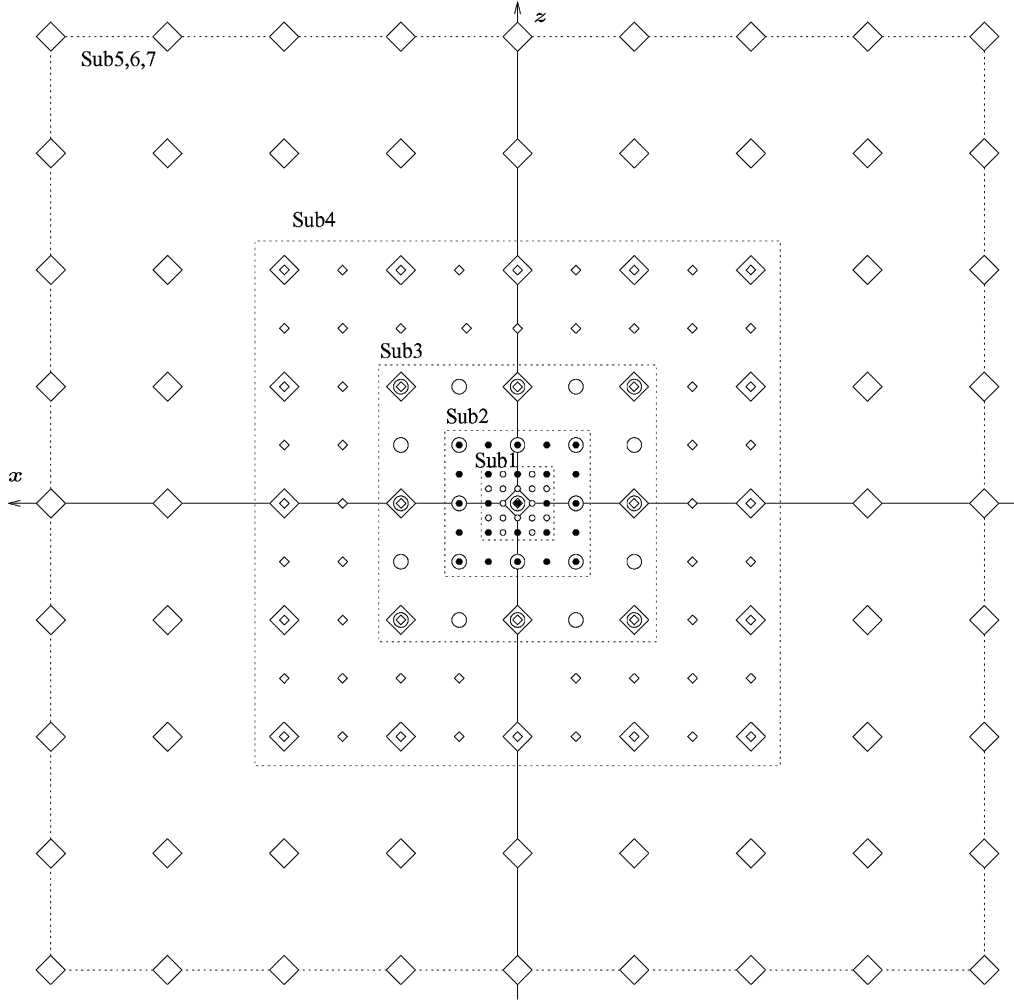
Fig. 2. Geometry of the virtual array generated by the four array sets shown in Fig. 1. Subarrays are harmonically nested to cover the acoustic band of $[50, 7200]$ Hz. The subarrays Sub1 to Sub3 each have 5×5 elements with $\lambda/2$ spacing; the subarrays Sub4 and Sub5 each have 9×9 elements with $\lambda/4$ spacing. Sub6 and Sub7 share the elements of Sub5. The size of the virtual array is 1.56×1.56 m^2.

is obtained by shifting the original arrays on the four corners toward the coordinate center and aligning them on the x axis and z axis, as illustrated in Fig. 3. The synthesized signal $\hat{S}_{0,0}(t)$ is the weighted average of the four delayed signals $S_{2,2}^{(1)}(t)$, $S_{2,-2}^{(2)}(t)$, $S_{-2,2}^{(3)}(t)$, and $S_{-2,-2}^{(4)}(t)$. The synthesized signal of a virtual element on the x axis or z axis is the weighted averages of two delayed signals. For example

$$\hat{S}_{0,-3}(t) = \frac{b_{0,-3}^{(1)}}{2} S_{2,-1}^{(1)} \left(t - \frac{\hat{r}_{0,-3} - r_{2,-1}^{(1)}}{c} \right)$$
$$+ \frac{b_{0,-3}^{(3)}}{2} S_{-2,-1}^{(3)} \left(t - \frac{\hat{r}_{0,-3} - r_{-2,-1}^{(3)}}{c} \right) \quad (2)$$

$$\hat{S}_{1,0}(t) = \frac{b_{1,0}^{(3)}}{2} S_{-1,2}^{(3)} \left(t - \frac{\hat{r}_{1,0} - r_{-1,2}^{(3)}}{c} \right)$$
$$+ \frac{b_{1,0}^{(4)}}{2} S_{-1,-2}^{(4)} \left(t - \frac{\hat{r}_{1,0} - r_{-1,-2}^{(4)}}{c} \right). \quad (3)$$

Other elements of the virtual array are synthesized by delaying the corresponding elements of the original array sets.

III. SUBBAND NEAR FIELD BEAMFORMING

A. Multirate Subband Processing

The synthesized signals of the virtual array are processed by the multirate subband beamformers shown in Fig. 4, which is similar to the subband scheme presented in our previous works [21], [27]. The input signals are first sampled at a high-frequency $F_s = 16$ kHz, then subbanded by an analysis filter $H_l(z)$ and decimated to a lower frequency F_l, where $l = 1, 2, \cdots, 7$. Noncritical sampling rate is used for each subband, that is $F_1 = F_s$, $F_{l+1} = F_l/2$ for $l = 1, 2, \cdots, 6$. Each subarray is followed by a broadband near-field beamformer using a tapped delay line designed for the corresponding subband. The outputs of the beamformers are interpolated and combined via the synthesis filters $G_l(z)$. The use of the multirate subband processing results in the same normalized frequency pass band for every subarray except Sub7. To be specific, $B_l = [0.225, 0.45]$ for $l = 1, \cdots, 6$ and $B_7 = [0.2, 0.45]$. However, to focus on a fixed near-field location, each subband beamformer has to be designed individually. This is due to the fact that the size of
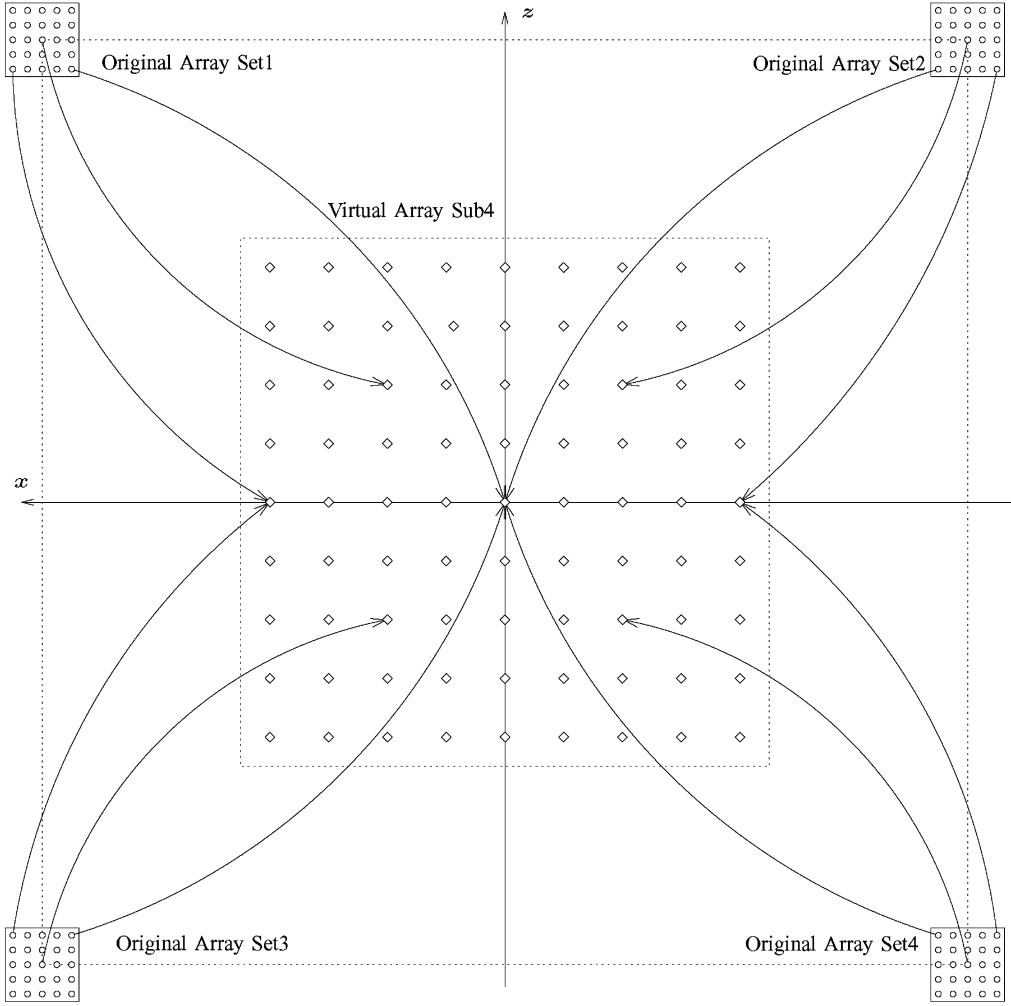
Fig. 3. Synthesizing the virtual array by the original array sets of Fig. 1.

each subarray and the radial distance of the focal point are different in terms of the corresponding wavelength. The details of the near-field beamformer design will be given in Section III-B.

The advantage of the nested array multirate subband beamforming technique is its reduced complexity. With multirate subband processing, the high-to-low frequency ratio of each subband reduces to 2:1. Therefore, the number of taps in each subband beamformer can be reduced substantially compared to a full-band beamformer. Nonuniform nesting of the subarrays also reduces the number of active elements in the virtual array in comparison to a uniform sampling scheme, because half-wavelength sampling at the highest frequency is grossly over sampled for lower frequencies. Therefore, nested array subband beamforming can reduce system complexity without performance loss.

B. Near-Field Beamformer Design

The near-field broadband beamformers are designed by optimization using the linearly constrained minimum variance (LCMV) method. Let the number of elements of the beamformer be M and the number of taps per element be K in the tapped-delay line filter. We have $N = MK$ degrees of freedom

for the beamformer optimization. The beamformer weights are determined by

$$\min_{\mathbf{w}}\{\mathbf{w}^{\dagger}\mathbf{R}\mathbf{w}\} \qquad (4)$$

$$\text{subject to} \qquad \mathbf{C}^{\dagger}\mathbf{w} = \mathbf{h} \qquad (5)$$

where the superscript $(\cdot)^{\dagger}$ denotes the Hermitian transpose, $\mathbf{w}$ is the concatenated weight vector, $\mathbf{R}$ is the $N \times N$ covariance matrix of the input signals, $\mathbf{C}$ is the constraint matrix, and $\mathbf{h}$ is the unit gain response vector. If the dimension of $\mathbf{C}$ is $N \times P$, then the constraint (5) is a set of P linear equations controlling the beamformer response. The constraints $\mathbf{C}$ and $\mathbf{h}$ are designed to enforce a unit gain at the desired signal location and over the desired temporal pass band. This is achieved by a small number of constraints with the near-field point constraint method [23]. Beamformer weights $\mathbf{w}$ are then optimized by minimizing the noise output power under a specific noise field.

Now, we propose a new noise model for the optimization design of the near-field broadband beamformers. The new noise model assumes that a large number of independent random noises are uniformly distributed over a sphere with finite radius
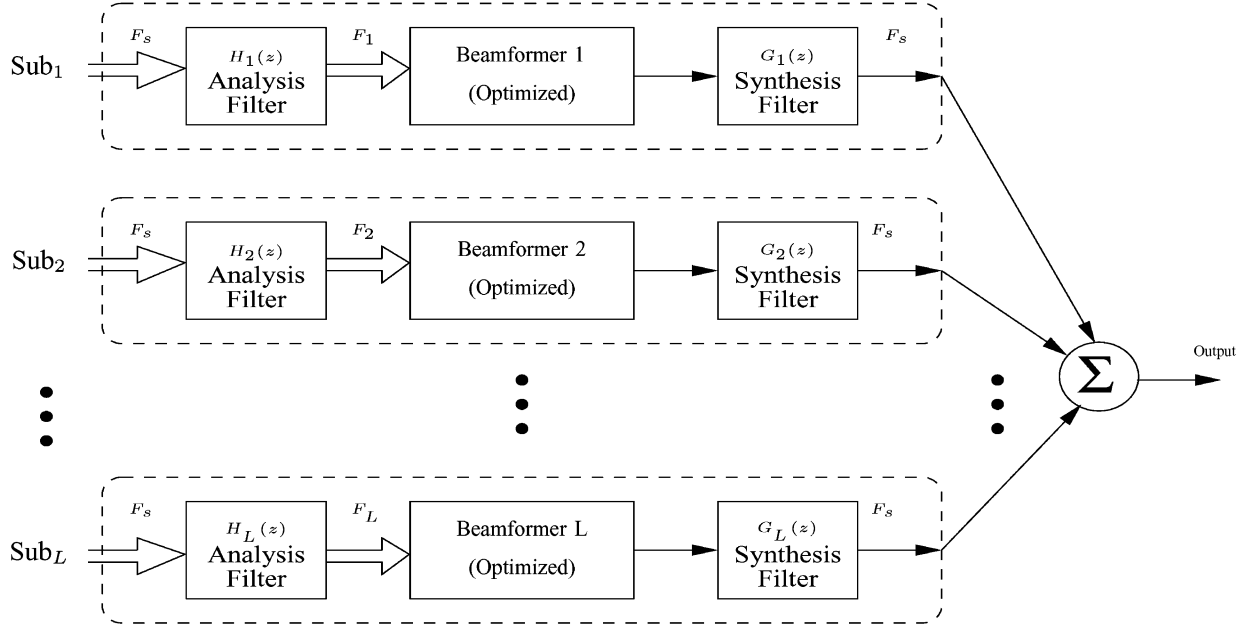
Fig. 4.   Subband multirate beamforming scheme for the nested virtual array. The sampling frequencies are: $F_s = 16$ kHz, $F_1 = F_s, F_{l+1} = F_l/2$ for $l = 1, 2, \cdots, L - 1$. Near-field beamformers are designed for each subband using constrained optimization under near-field spherically isotropic noise field.

r_n. The covariance matrix of the noise field observed at the sensor array is then

$$\mathbf{Q}_n = \frac{1}{4\pi B} \int_0^{2\pi} \int_0^\pi \int_B S_n(f)\mathbf{a}(\mathbf{x}_n, f)\mathbf{a}^\dagger(\mathbf{x}_n, f) \cdot \sin(\phi)df\,d\phi\,d\theta \quad (6)$$

where B is the normalized frequency band, $S_n(f)$ is the power spectrum density of the noises, and $\mathbf{a}(\mathbf{x}_n, f)$ is the near-field steering vector defined by

$$\mathbf{a}(\mathbf{x}_n, f) = \frac{r_n}{e^{j2\pi f r_n/c}} \left[\frac{e^{j2\pi f r_1/c}}{r_1}, \ldots, \right.$$
$$\left. \frac{e^{j2\pi f(r_m/c-k)}}{r_m}, \ldots, \frac{e^{j2\pi f(r_M/c-K+1)}}{r_M} \right]^T \quad (7)$$

and $r_m = |\mathbf{x}_m - \mathbf{x}_n|$ is the distance from the noise source location $\mathbf{x}_n = (r_n, \theta, \phi)$ to the mth element of the array located at $\mathbf{x}_m$.

The covariance matrix in (4) is chosen as

$$\mathbf{R} = \alpha\mathbf{Q}_n + \gamma\mathbf{I} \quad (8)$$

where $\mathbf{I}$ is the identity matrix, α and γ are the powers of the near-field isotropic noise field and the background noises, respectively. The noises are assumed to be white and having a flat power spectrum within the band of interests. The radius r_n and the powers α and γ are three design parameters used to trade off the beamformer robustness for noise suppression. The solution to the constrained optimization problem (4) and (5) is well known [25] as

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{C}(\mathbf{C}^\dagger\mathbf{R}^{-1}\mathbf{C})^{-1}\mathbf{h}. \quad (9)$$

The optimized weight vector will remain unchanged during the operation of the beamformer. Therefore, it will not have the desired signal cancellation problem generally encountered

by dynamically adaptive beamformers in reverberant environments.

The robustness of the designed beamformer is measured by its white noise gain defined [24] as $G_w = 1/(\mathbf{w}^\dagger\mathbf{w})$. The larger the white noise gain, the better robustness of the beamformer against background noises and errors. If the white noise gain G_w is too small, then the designed weight vector $\mathbf{w}$ has too large a norm and the beamformer output will have a poor signal-to-interference-and-noise ratio (SINR) due to the effect of background noise enhancement [25].

The novelty of the proposed optimization method is the use of the broadband, near-field, spherically isotropic noise model. Optimization under the far-field spherically isotropic noise model has been reported in the literature [19]. It is more convenient than the commonly used image model [26] which is dependent on the physical sizes and characteristics of the environment. A narrow-band near-field spherically isotropic noise model has also been used for frequency domain beamforming algorithms [17]. It shows improved performance over the far-field isotropic noise model when a very small radius of the noise sphere is chosen. However, reverberant signals are broadband in nature and are virtually originated from farther distances than the target signal due to reflections [26]. The proposed broadband, near-field, spherically isotropic noise model provides a direct form of reverberation modeling for broadband time domain beamforming. Using the new broadband near-field isotropic noise model, the performance of the near-field beamformer can be improved significantly when the radius of the noise sphere is larger than the distance between the array and the target signal. Furthermore, the radius of the near-field noise sphere provides a design parameter in addition to the power of the noise field for tradeoffs between de-reverberation gain and array robustness. It will be shown in Section IV that near-field beamformers designed by the new noise model achieve better de-reverberation gain than those designed
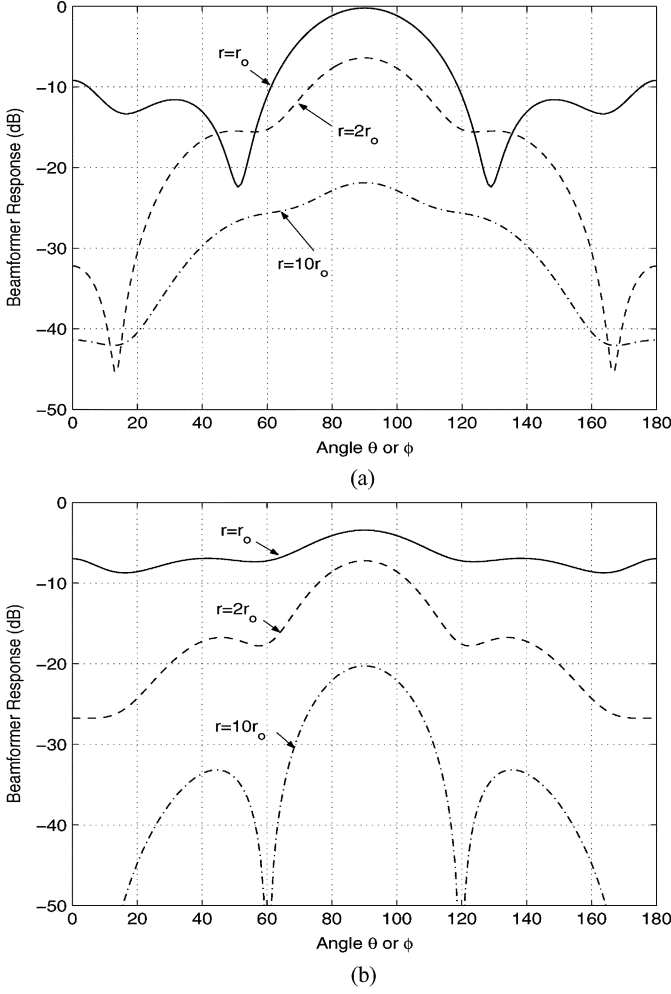
(a)

(b)

Fig. 5. Performance of the near-field beamformer compared to that of the far-field beamformer. All beampatterns are obtained by the virtual subarray Sub5 which has 9×9 elements equi-spaced at quarter wavelength of 450 Hz. Weights are optimized under white noises. The array responses are evaluated with $f = 400$ Hz and $r_o = 0.96$ m. Propagation attenuations are included.
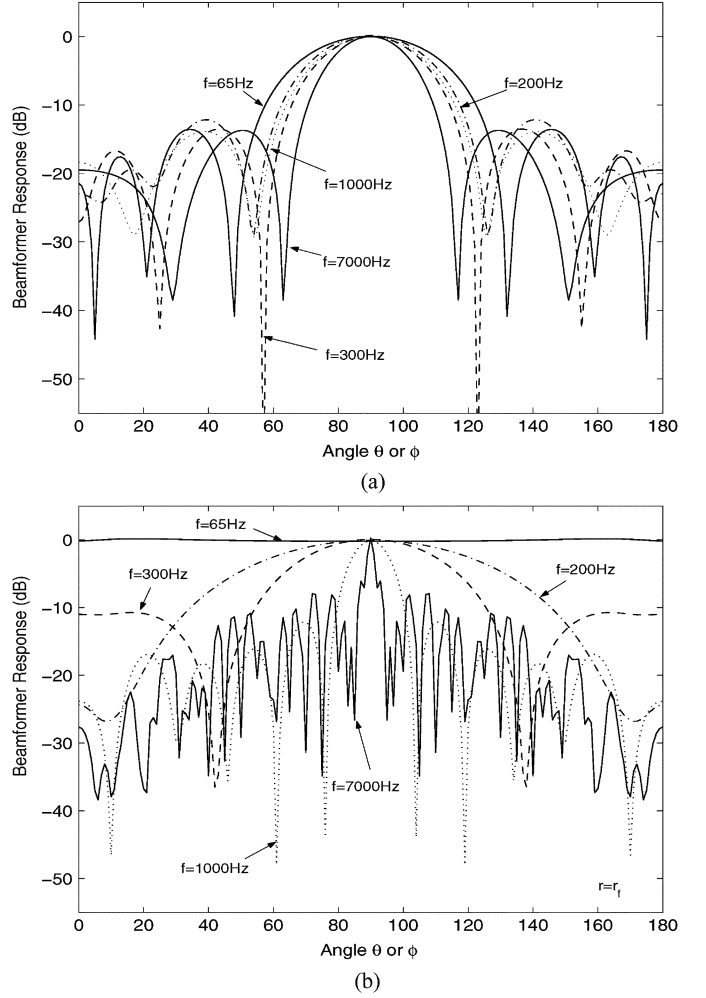

(a)

(b)

Fig. 6. Mainlobe beam width obtained by the subband beamformer compared to those of the full-band beamformer. Both beamformers use the same nested virtual array and near-field beamforming techniques with the focal point at $\mathbf{x}_o = (0.96 \text{ m}, 90°, 90°)$.

by the far-field isotropic noise model. By carefully choosing the radius and power of the noise sphere, the white noise gain can be kept within satisfactory level even for the super-gain beamformers of the low-frequency subbands.

IV. PERFORMANCES

This section presents the performances of the proposed near-field microphone array beamformer system evaluated via both computer simulations and real room experiments. First, the array directivity of the proposed multirate subband near-field beamforming method is compared to that of far-field beamforming and the full-band near-field field beamforming. Then, the proposed beamformer is evaluated in reverberant environments for its de-reverberation gain. Its robustness against errors is presented last.

A. Array Directivity

The array directivity of near-field beamforming is compared to that of conventional far-field beamforming using the example of subarray Sub5. Both beamformers had 9×9 active elements

and $K = 5$ taps per element. They were designed by the optimization method of (9) with $\alpha = 0$ and $\gamma = 0.01$. The near-field beamformer was focused at $\mathbf{x}_o = (0.96 \text{ m}, 90°, 90°)$. The far-field beamformer had a look direction at $(90°, 90°)$ without distance discrimination. Fig. 5 shows the beampatterns obtained with simulation at 400 Hz and three radial distances from the array center. Fig. 5(a) shows that the near-field beamformer provided good directivity at the focal point $(r = r_o)$ while attenuating 10 dB or more at sidelobes and far away locations. On the other hand, Fig. 5(b) shows that the far-field beamformer demonstrated little spatial directivity in near-field areas at distances $r = r_o$ and $r = 2r_o$. Instead, good directivity was exhibited at the far away distance of $r \geq 10r_o$. Performance degradation by far-field beamforming is obvious over the near-field target region, and it is more severe at low frequencies. Consequently, far-field beamforming is not suitable for applications with signal targets located in the near field of the array.

The advantage of the multirate subband beamforming is illustrated in terms of reduced beampattern variations, as shown in Fig. 6. The mainlobe beam widths of the subband beamformer
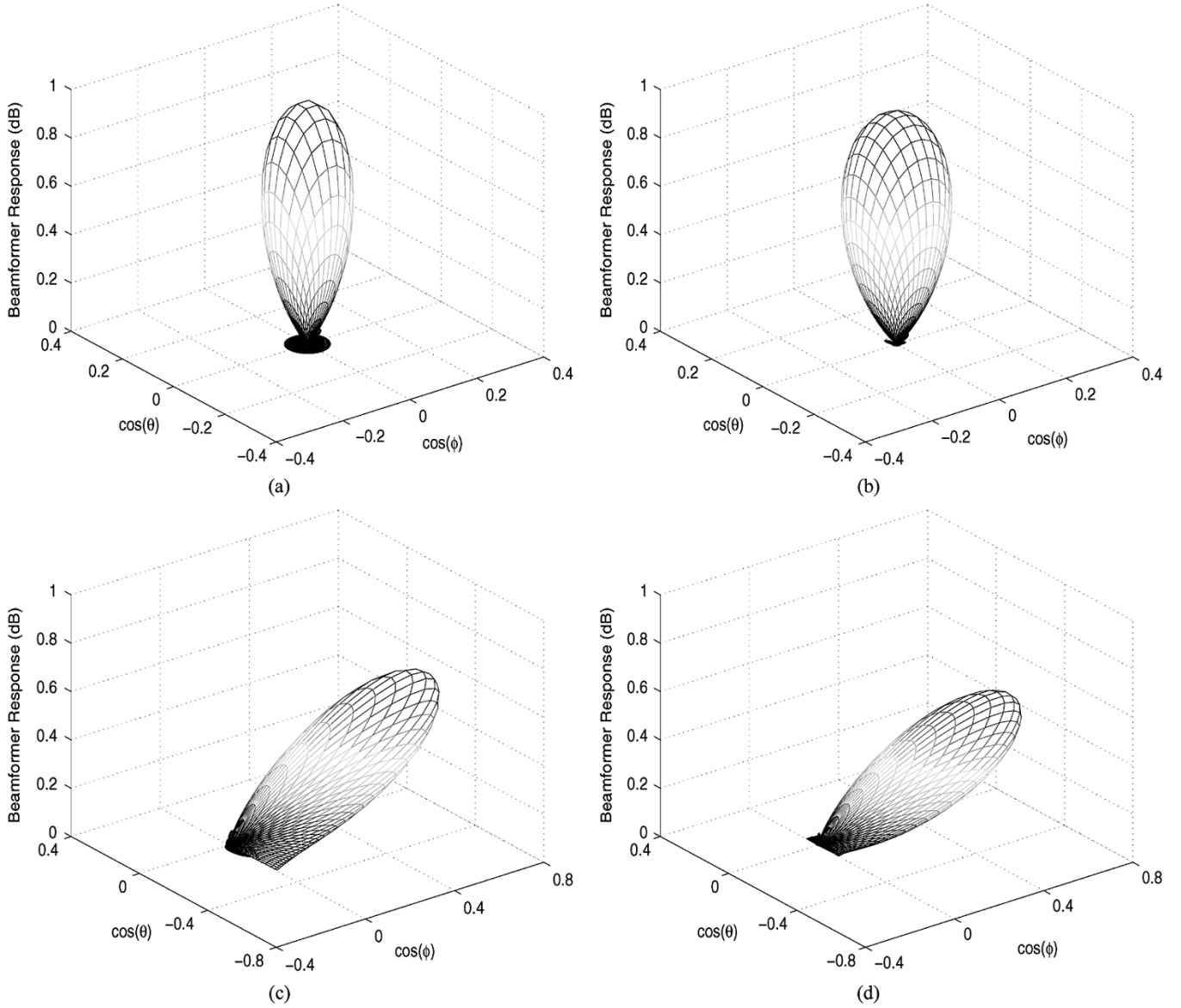
Fig. 7.   Polar beampatterns obtained by the subband nested array beamformer evaluated at the frequency f and the focal point $\mathbf{x}_o$. (a) $f = 6000$ Hz, $\mathbf{x}_o = (0.96\,\text{m}, 90°, 90°)$, (b) $f = 500$ Hz, $\mathbf{x}_o = (0.96\,\text{m}, 90°, 90°)$, (c) $f = 6000$ Hz, $\mathbf{x}_o = (0.96\,\text{m}, 120°, 50°)$, and (d) $f = 500$ Hz, $\mathbf{x}_o = (0.96\,\text{m}, 120°, 50°)$.

are compared to those of the full-band beamformer. Both beamformers used the same nested virtual array structure of Fig. 2 and the same near-field optimization method proposed in Section III-B with $\alpha = 10$ and $\gamma = 0.01$ and the focal point $\mathbf{x}_o = (0.96\,\text{m}, 90°, 90°)$. The radius of the noise field r_n was selected within the range of $3r_o$ to $10r_o$. The subband beamformer used five taps per element while the full-band beamformer used 11. The beampatterns were evaluated at $r = r_o$ and several in-band frequencies. The mainlobe beam width of the full-band beamformer, as shown in Fig. 6(b), widens as the frequency decreases. The beam width at low frequencies is too large to provide adequate directivity. In contrast, the nested array subband beamformer reduced the 3-dB mainlobe beam width variations to within $15°$, as shown in Fig. 6(a). This is satisfactory in most applications.

Fig. 7 illustrates the three-dimensional (3-D) polar beampatterns of the subband near-field beamformers. With the focal point at $\mathbf{x}_o = (0.96\,\text{m}, 90°, 90°)$ and $\mathbf{x}_o = (0.96\,\text{m}, 120°, 50°)$,

respectively, the subband near-field beamformers obtained similar beampatterns at different frequencies. The beam width variations are very small.

B. De-Reverberation Performance

We now show the de-reverberation performance of the proposed near-field subband beamformers by both simulations and real room experiments. In the simulation, the impulse response of a reverberant room was generated by the image model [26]. The simulated room had a size of $5.0 \times 3.8 \times 3.5$ m^3 with the reflection coefficients of the walls being 0.9 and those of the ceiling and floor being 0.7. The reverberation time of the simulated room was approximately $T_{60} \approx 300$ ms, which is typical for office rooms. The four planar array sets were located in a corner of the room with its phase center at 1.0 m away from the floor and the walls. The array plane was perpendicular to the floor and at $45°$ angles with the walls, as shown in Fig. 8. An audio signal source was located in front of the array on the y
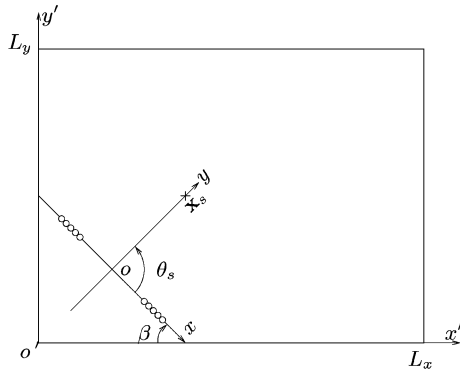
Fig. 8. Simulated reverberant room. The reverberation time of the room is $T_{60} \approx 300$ ms. The angle between the array axis and the wall is $\beta = 45°$. The signal source is located at $\mathbf{x}_s = (1.0 \text{ m}, 90°, 90°)$ in the array coordinates. The figure is not to scale.

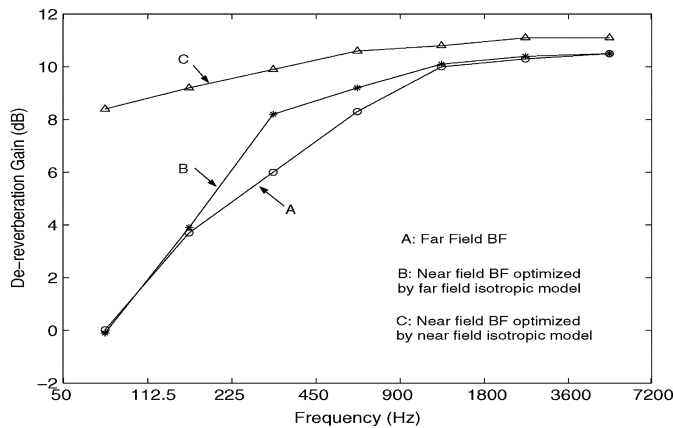Fig. 10. Objective mean opinion scores obtained by the PESQ test using a DSLA.

Fig. 9. De-reverberation performances of the nested array subband beamformers. All curves were obtained by the nested virtual array under the simulated reverberation environment shown in Fig. 8. The uncorrelated background white noise was 20 dB below the desired signal for all beamformers.

axis with a distance of 1.0 m. The reverberant noises were generated by convolution of the clean signal with the room impulse responses.

The signal and the reverberant noises were processed by the nested array subband beamformers. The de-reverberation performances were measured by the output SINRs of the beamformers computed for each subband, as plotted in Fig. 9. The SINR of the proposed near-field beamformer (curve C) is compared to those of the far-field beamformer (curve A) and the conventional near-field beamformer optimized by far-field isotropic noise model (curve B). The conventional near-field beamformer improved its de-reverberation gain by 1 or 2 dB over the far-field beamformer only in the two middle-frequency subbands. For the high- and low-frequency subbands, it had almost no improvement over the far-field beamformer. In comparison, the proposed near-field beamformer achieved better de-reverberation gain than that of the far-field beamformer over all subbands: close to 1-dB improvement over the three high-frequency subbands, 3 to 5 dB over the middle bands, and 8 dB on the lowest subband. This significant improvement is due to the fact that the low- and middle-frequency subarrays have larger near-field regions than high-frequency ones and the near-field isotropic noise model helped to suppress the near-field noises. Based on
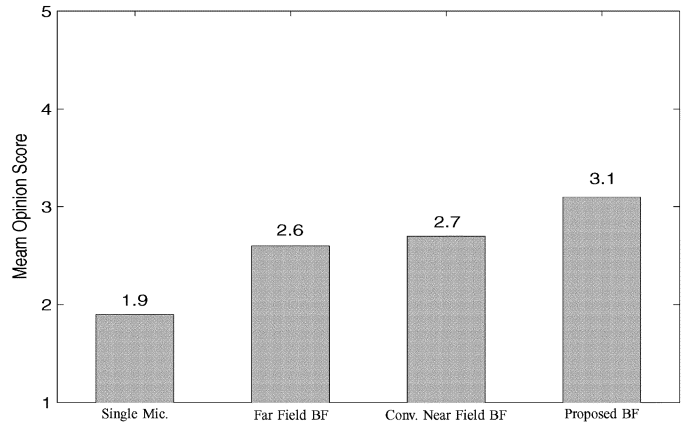
the image model [26], a large number of reflected signals were within 10 m from the center of the array. The closest image source was located at a distance of 2.2 m. This distance is well within the near field of subarrays Sub4 to Sub7 and not far from the near-far boundary of Sub1 to Sub3. Thus, the reverberant noises are better modeled by the proposed near-field spherically isotropic noise model than the conventional far-field isotropic model, resulting in better performances.

Real room experiments were also performed to verify the de-reverberation performance of the proposed array system. The experiment was conducted in a small conference room with the reverberation time being more than 300 ms. The room settings, recording equipment, and furniture arrangement were the same as those described in [27]. The background noise level was kept low so that it was particularly suitable for evaluating the de-reverberation performance. The four microphone array sets were placed on a desk in a corner of the room with the center of the sets located at 1.0 m away from the walls and the floor. The array sets and sound sources were arranged similar to the settings in the computer simulation, as shown in Fig. 8. Fifteen clean speech signals were used and recorded by the microphone arrays. The data were then processed by the three beamformers.

A PESQ test was performed using a digital speech level analyzer (DSLA) made by Malden Electronics, Ltd. The PESQ provides an objective measure of speech quality defined in ITU-T Recommendation P.862/2001. It uses a sensory model to compare the original (reference) signal to the degraded version and the result of the comparison is a quality score. The PESQ score is analogous to the subjective MOS determined by a well controlled subjective test according to ITU-T Recommendation P.800 [28]. In our experiment, the input signals of a single microphone and the output signals of the beamformers were compared to the clean signal sources and the objective MOS scores were averaged over the fifteen speech signals. The results are shown in Fig. 10, where the single microphone inputs scored the lowest MOS of 1.9 points since these signals contained the highest reverberant interference. The far-field beamformer obtained 2.6 points because the beamformer's directivity eliminated a large number of reverberant noises. The conventional near-field beamformer scored 2.7 points which was similar to the far-field beamformer. The proposed near-field beamformer

TABLE I
WHITE NOISE GAIN OF THE SUBBAND BEAMFORMERS

	Far Field BF	Conv. Near Field BF	Proposed NF BF
50–112.5 Hz	21.3 dB	17.2 dB	−10.6 dB
112.5–225 Hz	21.3 dB	17.1 dB	−8.7 dB
225–450 Hz	21.3 dB	17.0 dB	−2.0 dB
450–900 Hz	21.3 dB	18.2 dB	−2.6 dB
900–1800 Hz	16.2 dB	13.7 dB	0.9 dB
1.8–3.6 kHz	16.2 dB	13.9 dB	1.0 dB
3.6–7.2 kHz	16.2 dB	14.0 dB	1.0 dB

obtained 3.1 points due to its more than 8 dB de-reverberation gain over all subbands. The proposed beamformer provided 1.2 points improvement over the reverberant inputs and more than half a MOS point over the two conventional beamformers. This improvement in speech quality is significant and it is primarily due to its enhanced de-reverberation gain in low-frequency subbands.

C. Robustness Against Errors

The performance improvement of the proposed near-field beamformer over the other near/far-field beamformers was obtained at the cost of slightly reduced white noise gain. As shown in Table I, the far-field beamformer and the conventional near-field beamformer had larger white noise gains meaning good robustness against background noises. However, for the background noise at the level of 20 dB below the desired signal, a white noise gain of −15 dB is sufficient. In the proposed design method, the powers α and γ, and the radius r_n can be used to trade off between the de-reverberation gain and white noise gain. The parameters used in our design were $\alpha = 10, \gamma = 0.01$, and $3r_o < r_n < 10r_o$. The white noise gains of the resulting beamformers were larger than −11 dB and their robustness was satisfactory.

Robustness against the signal location and synthesis error was also important because the location estimation is difficult and often not accurate for near-field signal sources [4]. Beamformers have to tolerate small location errors of the desired signal while maintaining large attenuations to signals from other locations [23]. This property was evaluated by moving the desired signal away from the presumed location and computing the array gain at the actual signal location. Let the signal location be $\mathbf{x} = (r, \theta, \phi)$ and the presumed signal location or the focal point of the beamformers be $\mathbf{x}_o = (0.96, 90°, 90°)$. When the signal location had distance errors ($r \neq r_o$) or Angle of Arrival (AoA) errors ($\Delta\theta \neq 0$, or $\Delta\phi \neq 0$), the array gains were reduced and the average gain losses are shown in Fig. 11(a) and (b), respectively. The gain reduction due to distance errors was less than 1 dB when the signal was located within the range of $[0.7r_o, 2r_o]$. A small distance error toward the array ($r/r_o < 1$) caused the gain to decrease sharply. While the signal was moved away from the presumed location ($r/r_o > 1$), a slower gain loss was exhibited. This is because the near-field beamformers have better distance discrimination in the near-field region than in the far-field region. Meanwhile, Fig. 11(b) shows that the beamformer was more sensitive to the AoA errors. The gain was reduced by 1 dB when the AoA error was ±5° and the 3-dB gain loss occurred when the AoA error was larger than ±20°. Considering both the distance and AoA errors, the 1-dB gain loss re-
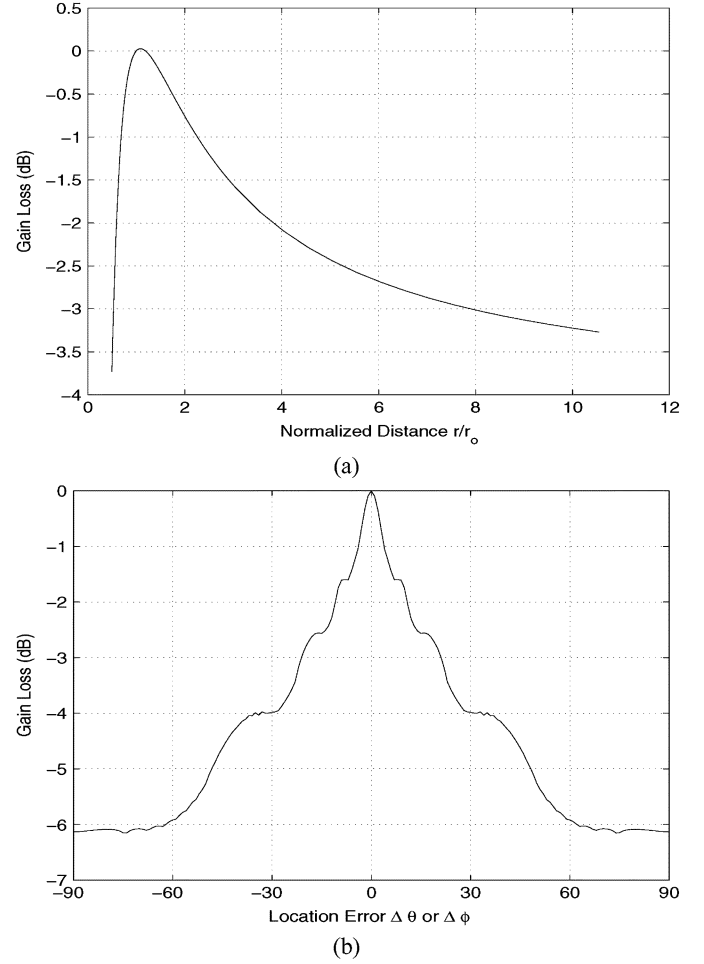

Fig. 11. Robustness against location errors measured by the average gain loss with respect to the array gains without location errors. (a) Location error in distance while the AoA has no error. (b) Location error in AoA while the distance has no error.

gion was a 3-D region with $r = [0.7, 1.9]$ meter, $\theta = [85°, 95°]$, and $\phi = [85°, 95°]$. These angles were approximately equal to ±10 cm in sideways and in height around the presumed focal point. The overall sensitivity to signal location errors was satisfactory according to the accuracy of location estimation.

V. CONCLUSION

A microphone array beamforming system has been proposed for multimedia communication applications, where four sets of small-sized 5×5 planar arrays have been used on a computer monitor. Two new techniques have been proposed for the array beamforming system to optimize the performances in reverberant environments. One is the virtual array approach which synthesizes a nonuniformly spaced nested virtual array from the original array sets by weighted average of the delayed original signals to cover the frequency band of [50, 7000] Hz for the wideband telephony. The second new technique is a new noise model, namely the broadband near-field spherically isotropic noise model, for optimization of the subband near-field beamformers. The new noise model employs a large number of independent random noises uniformly distributed over a sphere with a finite radius. The radius of the noise model serves as

a design parameter in addition to the noise power for trade-offs between performance and robustness. Computer simulations and real room experiments have shown that the proposed array beamforming system has improved the performances of sound qualities while maintaining sufficient robustness against white noises and location errors.

REFERENCES

[1] H. Wang and L. Y. Wang, "Multi-sensor adaptive heart and lung sound extraction," in *Proc. IEEE Sensors Conf.*, Toronto, ON, Canada, Oct. 2003, pp. 1096–1099.

[2] Q. Lin, E. E. Jan, and J. Flanagan, "Microphone arrays and speaker identification," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 5, pp. 622–629, Sep. 1994.

[3] A. A. Handzel and P. S. Krishnaprasad, "Biomimetic sound-source localization," *IEEE Sensors J.*, vol. 2, no. 6, pp. 607–616, Dec. 2002.

[4] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1843–1854, Aug. 2002.

[5] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 3, pp. 240–259, May 1998.

[6] M. Dahl and I. Claesson, "Acoustic noise and echo canceling with microphone array," *IEEE Trans. Veh. Technol.*, vol. 48, no. 5, pp. 1518–1526, Sep. 1999.

[7] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: Approaches, applications, and future research directions," *IEEE Sensors J.*, vol. 2, no. 2, pp. 107–119, Apr. 2002.

[8] D. Zotkin, R. Duraiswami, V. Philomin, and L. S. Davis, "Smart Video-conferencing," in *Proc. IEEE Int. Conf. Multimedia Expo.*, vol. 3, ICME-2000, pp. 1597–1600.

[9] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West, and M. M. Sondi, "Autodirective microphone systems," *Acustica*, vol. 73, pp. 58–71, 1991.

[10] F. Khalil, J. P. Jullien, and A. Gilloire, "Microphone array for sound pickup in teleconference systems," *J. Audio Eng. Soc.*, vol. 42, no. 9, pp. 691–700, Sep. 1994.

[11] M. S. Branstein and D. B. Ward, "Cell-based beamforming (CE-BABE) for speech acquisition with microphone arrays," in *IEEE Trans. Speech Audio Process.*, vol. 8, Nov. 2000, pp. 738–743.

[12] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," in *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, Jun. 1986, pp. 393–398.

[13] W. Täger, "Near field superdirectivity (NFSD)," in *Proc. IEEE ICASSP*, vol. 4, May 1998, pp. 2045–2048.

[14] S. Chowdhury, M. Ahmadi, and W. C. Miller, "Design of a MEMS acoustical beamforming sensor microarray," *IEEE Sensors J.*, vol. 2, no. 6, pp. 617–627, Dec. 2002.

[15] D. P. Arnold, T. Nishida, L. Cattafesta, and M. Sheplak, "MEMS-based acoustic array technology," *J. Acoust. Soc. Amer.*, vol. 113, no. 1, pp. 289–298, Jan. 2003.

[16] J. G. Ryan, "Criterion for the minimum source distance at which plane-wave beamforming can be applied," *J. Acoust. Soc. Amer.*, vol. 104, no. 1, pp. 595–598, Jul. 1998.

[17] T. D. Abhayapala, R. A. Kennedy, and R. C. Williamson, "Noise modeling for nearfield array optimization," *Electron. Lett.*, vol. 35, no. 10, pp. 764–765, May 13, 1999.

[18] ——, "Spatial aliasing for near-field sensor arrays," *Electron. Lett.*, vol. 35, no. 10, pp. 764–765, May 13, 1999.

[19] J. G. Ryan and R. A. Goubran, "Array optimization applied in the near field of a microphone array," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 173–176, Mar. 2000.

[20] ——, "Application of near-field optimum microphone arrays to hands-free mobile telephony," *IEEE Trans. Veh. Technol.*, vol. 52, no. 2, pp. 390–400, Mar. 2003.

[21] Y. R. Zheng, R. A. Goubran, and M. El-Tanany, "A nested sensor array focusing on near field targets," in *IEEE Int. Conf. Sensors*, Toronto, ON, Canada, Oct. 2003, pp. 843–848.

[22] P. Mermelstein, "G.722, a new CCITT coding standard for digital transmission of wideband audio signals," *IEEE Commun. Mag.*, vol. 26, no. 1, pp. 8–15, Jan. 1988.

[23] Y. R. Zheng, R. A. Goubran, and M. El-Tanany, Robust near field adaptive beamforming with distance discrimination, in *IEEE Trans. Speech Audio Process.*, to be published.

[24] H. Cox, R. M. Zeskind, and M. H. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[25] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, pp. 4–24, Apr. 1988.

[26] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[27] Y. R. Zheng, R. A. Goubran, and M. El-Tanany, "Experimental evaluation of a nested microphone array with adaptive noise cancellers," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 3, pp. 777–786, Jun. 2004.

[28] Digital Speech Level Analyzer User Guide (2004). [Online]. Available: http://www.malden.co.uk/downloads.htm

Yahong Rosa Zheng (S'99–M'03) received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, in 1987, the M.S. degree from Tsinghua University, Beijing, China, in 1989, and the Ph.D. degree from Carleton University, Ottawa, ON, Canada, in 2002, all in electrical engineering.

From 1989 to 1997, she was with several companies, including the Peony Electronic Group, Beijing, China; Sagem Australasia Pty., Ltd., Sydney, Australia; and Polytronics Pty., Ltd., Toronto, ON, Canada. From 2003 to 2005, she was an NSERC Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Missouri, Columbia. Currently, she is an Assistant Professor with the Department of Electrical and Computer Engineering, University of Missouri, Rolla. Her research interests include array signal processing, wireless communications, and wireless sensor networks.

Dr. Zheng has served as a TPC member for several IEEE international conferences including ICC, GlobeCom, and the Sensors Conference.

Rafik A. Goubran (M'85) received the B.Sc. and M.Sc. degrees in electrical engineering from Cairo University, Cairo, Egypt, in 1978 and 1981, respectively, and the Ph.D. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 1986.

In January 1987, he joined the Department of Systems and Computer Engineering, Carleton University, where he is now a Professor and Chair. He was involved in several research projects with industry and government organizations, including Nortel Networks, Mitel Networks, Bell Canada, NEC Corporation, the Department of National Defense, and the National Research Council of Canada. His research interests include digital signal processing and its applications in audio and biomedical engineering, voice transmission over IP networks, noise and echo cancellation, and beamforming using microphone arrays.

Mohammed El-Tanany received the B.Sc. and M.Sc. degrees in electrical engineering from Cairo University, Giza, Egypt, in 1974 and 1978, respectively, and the Ph.D. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 1983.

He was with the Advanced Systems division of Miller Communications, Kanata, ON, from 1982 to 1985, with principal involvement in the research and development of digital transmission equipment for mobile satellite-type of applications and also for VHF airborne high-speed down links. He joined Carleton University in 1985, where he is currently a Professor with the Department of Systems and Computer Engineering. His research activities are mainly in the areas of digital wireless communications with emphasis on the transmission subsystems from a hardware design and algorithm design points of view, digital TV and digital audio broadcasting systems, experimental characterization and empirical modeling of the wireless communications channels, and digital TV channels in various environments and at various frequency bands.

Hongchi Shi (M'95–SM'99) received his Ph.D. degree in computer and information sciences from the University of Florida, Gainesville, in 1994.

He is currently an Associate Professor with the Department of Computer Science, University of Missouri–Columbia. His research interests include parallel and distributed computing, wireless sensor networks, image processing, and neural networks. He has served on the program committees of several international conferences. He has chaired the SPIE annual conference on parallel and distributed methods for image processing for several years.