Microphone array signal processing with application in three-dimensional spatial hearing

Mingsian R. Bai[a] and Chenpang Lin
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan, Republic of China

(Received 3 August 2004; revised 4 December 2004; accepted 6 December 2004)

Microphone arrays are known to enhance the directionality and signal-to-noise ratio (SNR) over single-channel sensors. This is considered beneficial in many applications such as video-conferencing systems and hearing aids. However, this advantage comes at the price of the sensation of spatial hearing. The spatial cues due to diffractions of the head and torso are lost if the array is not fitted in the ears. In this paper we present a system that incorporates binaural hearing synthesis into array signal processing, in an attempt to recover the three-dimensional sound image that a human listener would naturally perceive. In the system, the superdirective beamformer is exploited to estimate the direction of arrival (DOA) of the incoming sound. The spatial sound image is restored by steering the beam to the direction found in the DOA session and filtering the array output with the corresponding Head Related Transfer Functions (HRTF). The algorithms have been implemented in real-time fashion using a digital signal processor. Objective and subjective experiments were performed to validate the proposed system. The experimental results showed that the accurate localization of the sound source is achievable using the array system. [DOI: 10.1121/1.1853242]

PACS numbers: 43.60.Fg [EJS] Pages: 2112–2121

I. INTRODUCTION

Microphone arrays have received much research interest as a means of acoustic pickup utilized in various applications such as video-conferencing systems[1] and hearing aids.[2–5] One of the reasons for using an array is to improve signal-to-noise ratio (SNR) that has long been a plaguing problem of, for example, conventional single channel hearing aids. This problem is further aggravated in the environments where reverberations and interferences are present. In comparison with single-channel sensors, microphone arrays provide advantages that the SNR as well as directionality of the sensor can be enhanced using such a system. In particular, an array behaves like a spatial filter, enabling the listener to focus on the signal source such as speech and at the same time reject ambient noise and interference.[6] This is an attractive feature for hearing-impaired people who may desire low-noise hearing aids. A great number of array signal processing methods can be found in the literature[7] to design arrays subject to individual requirements in an application. Beamforming and estimation of direction of arrival (DOA) are known to be two major functions of arrays. These two functions serve to track the intended source in a particular direction with high signal quality.

However, the above-mentioned advantages come at the price of sensation of spatial hearing. The spatial cues due to diffractions of the head and torso are lost if the sensors are not fitted in ears. To address the problem, in this paper we present a system that incorporates binaural hearing synthesis into array signal processing, in an attempt to recover the three-dimensional (3-D) sound image that a human listener would naturally perceive. In order to form a very sharp beam, the superdirective array[8–13] is employed in the array design. The thus designed beam is then electronically steered in every direction to estimate the DOA. Once the DOA is found, the sound beam is fixed at that direction. Finally, the binaural signals of the spatial sound image are produced by filtering the array output with the corresponding Head Related Transfer Functions (HRTF).[14] The algorithms are implemented in real-time fashion using a digital signal processor, TMS320C32. As such, a low noise sound with high spatial quality is reproduced with the aid of the thus integrated array–HRTF system.

Simulations and experiments are carried out to evaluate the proposed system. The SNR gain and the effects of aperture size on the directivity of the microphone arrays are examined. Objective and subjective tests were performed to see how effective the listeners localize the source of sound using the proposed array system.

II. MICROPHONE ARRAY SIGNAL PROCESSING

It is well known that microphone arrays are capable of enhancing the directionality and signal-to-noise ratio (SNR) over single-channel sensors. This is beneficial in many applications such as hearing aids. In this paper, array signal processing techniques are utilized for the DOA estimation and beamforming. DOA estimation refers to localizing the source direction, while beamforming refers to forming a beam pattern with a specified shape and orientation. In this section, two algorithms will be presented to carry out these tasks of array signal processing, followed by an investigation on the directivity pattern.

[a]Corresponding author. Electronic mail: msbai@mail.nctu.edu.tw

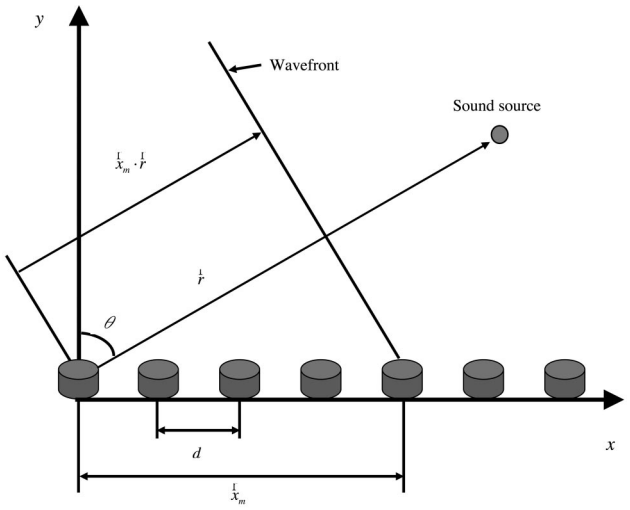


FIG. 1. The configuration of a uniform linear array (ULA). A sound source is located in the farfield. $\vec{x}_m$ is the position vector of the *m*th sensor and $\vec{r}$ is the position vector of the farfield source.

A. Array signal processing algorithms

Figure 1 shows a uniform linear array (ULA), in which sensors are allocated along a straight line with equal spacing *d*. It is assumed that the sound source is at the farfield and the signals received at the sensors are narrowband with center frequency ω. The spacing between adjacent microphones is less than one-half wavelength to avoid grating lobes.

1. The delay-sum array

A brief review of the delay-sum method is given as follows. For an *M*-sensor array, the signals received at the sensors, $x_1(t),\ldots,x_M(t)$, form the *data vector*

$$\mathbf{x}(t)=\mathbf{a}(\vec{r})s(t)+\mathbf{n}(t), \tag{1}$$

where $\mathbf{a}(\vec{r})=[e^{j\omega[(\vec{r}_1\cdot\vec{r})/c]}\cdots e^{j\omega[(\vec{r}_M\cdot\vec{r})/c]}]^T$ is termed the *array manifold vector*, $\vec{r}$ is the unit vector pointing to the source, $\vec{r}_m$, $m=1,\ldots,M$ are the position vectors of the sensors, *c* is the speed of sound, $s(t)$ represents the baseband signal of the source and $\mathbf{n}(t)=[n_1(t)\cdots n_M(t)]^T$ is the noise vector. The beamformer output is the weighted sum of the delayed input signals, given by

$$y(t)=\mathbf{w}^H\mathbf{x}(t)=\mathbf{w}^H\mathbf{a}(\vec{r})s(t)+\mathbf{w}^H\mathbf{n}(t)], \tag{2}$$

where $\mathbf{w}=[w_1\cdots w_M]^T$ is the array weight vector and the operator ^{H} denotes the complex conjugate transposition. The delay-sum algorithm is to ''time align'' the received signals by choosing the following weight vector:

$$w_m=\frac{1}{M}e^{-j\omega[(m-1)d\sin\theta/c]}=e^{-j\omega\tau_m},\quad m=1,\ldots,M, \tag{3}$$

where θ is the look angle (measured from the normal) of the array and is dependent of the source vector $\vec{r}$ and

$$\tau_m=\frac{(m-1)d\sin\theta}{c} \tag{4}$$

is the delay that the *m*th channel needs to compensate. It can be shown that the delay-sum algorithm attains the maximum signal-to-noise ratio gain (SNRG).[7] The delay τ_m usually is not an integer. The simplest approach to deal with these fractional delays is the Lagrange interpolation method.[15] We first divide τ_m by the sampling period *T* to acquire the fractional delay that can be written into the integer and fractional components, D_m and e_m, respectively,

$$\frac{\tau_m}{T}=D_m+e_m. \tag{5}$$

The Finite Impulse Response (FIR) filter coefficients required to realize the fractional delay are given by

$$w_{mk}=\prod_{\substack{l=0\\ l\neq k}}^{N}\frac{e_m-l}{k-l},\quad k=0,1,2,\ldots,N. \tag{6}$$

The coefficients for the Lagrange filters of order $N=1$, 2 are given in Table I. The expectation value of $|y(t)|^2$ plotted versus the look angle θ is called the spatial power spectrum:

$$S(\theta)=E\{|y(t)|^2\}=\mathbf{w}^H\mathbf{R}_{xx}\mathbf{w}, \tag{7}$$

where $\mathbf{R}_{xx}\triangleq E\{|\mathbf{x}(t)\mathbf{x}^H(t)|\}$ is the *data correlation matrix*. The peak of the spatial power spectrum corresponds to the direction of the sound source.

TABLE I. FIR filter coefficients for the first-order ($N=1$) and the second-order ($N=2$) Lagrange interpolation.

	w_{m0}	w_{m1}	w_{m2}
$N=1$	$1-e_m$	e_m	
$N=2$	$(e_m-1)(e_m-2)/2$	$-e_m(e_m-2)$	$e_m(e_m-1)/2$

2. The superdirective array

Another array algorithm employed in this work is the superdirective array.[13] The principle of this algorithms follows from maximizing the *array gain* that is the measure of improvement of the signal-to-noise ratio between one sensor and the array output,

$$G=\frac{\mathrm{SNR}_{\mathrm{Array}}}{\mathrm{SNR}_{\mathrm{Sensor}}}. \tag{8}$$

The larger the array gain, the higher the ability of the array as a spatial filter to suppress the noise. The above array gain can be shown to be equivalent to[13]

$$G=\frac{|\mathbf{w}^H\mathbf{d}|^2}{\mathbf{w}^H\mathbf{\Gamma}_{nn}\mathbf{w}}, \tag{9}$$

where $\mathbf{\Gamma}_{nn}$ is the *coherence matrix* of the noise and $\mathbf{d}$ is the steering vector of the main axis of the array. It is generally assumed that a diffuse white noise field is spherically isotropic, in which the *ab*th entry of the matrix takes the form[13]

$$\Gamma_{n_an_b}=\frac{\sin[k(a-b)d]}{k(a-b)d}. \tag{10}$$

On the other hand, if only the self-noise of the sensors is present, i.e., $\mathbf{\Gamma}_{nn}=\mathbf{I}$, the array gain *G* reduces to the white noise gain (WNG):

$$\mathrm{WNG}=\frac{|\mathbf{w}^H\mathbf{d}|^2}{\mathbf{w}^H\mathbf{w}}. \tag{11}$$

In general, the larger the array weights, the smaller the WNG and hence the more sensitive to noise is the array. Another important quantity to evaluate arrays, the directivity index (DI) is the logarithmic equivalent of the above-mentioned array gain,

$$DI = 10 \log_{10}\left(\frac{|\mathbf{w}^H \mathbf{d}|^2}{\mathbf{w}^H \mathbf{\Gamma}_{nn} \mathbf{w}} \right), \tag{12}$$

which amounts to the ratio of the main-axis gain over the angle-weighted gain. The larger the DI, the more directional is the array pattern.

The idea of the superdirective design is realized by maximizing the DI in Eq. (12), or equivalently, the array gain in Eq. (8). Equation (8) is in fact a Rayleigh's quotient. The maximization of which can be achieved by solving the following optimization problem:

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{\Gamma}_{nn} \mathbf{w}, \quad \text{subject to} \quad \mathbf{w}^H \mathbf{d} = 1. \tag{13}$$

In other words, we look for an optimal weight $\mathbf{w}$ such that the array output power is minimized with the gain at the look direction constrained to unity. Thus, the array aims to receive an undistorted signal response at the main axis and reject unwanted interferences at the other directions. Following the method of the Lagrange multiplier, the solution of Eq. (13) is given as[7,16]

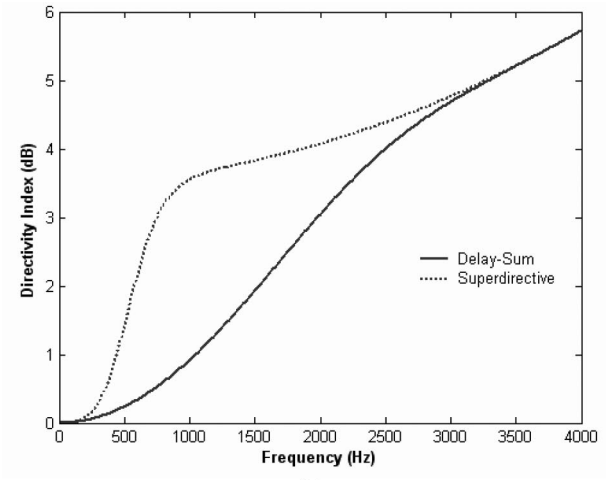
$$\mathbf{w} = \frac{\mathbf{\Gamma}_{nn}^{-1} \mathbf{d}}{\mathbf{d}^H \mathbf{\Gamma}_{nn}^{-1} \mathbf{d}}. \tag{14}$$

In the low-frequency range, the matrix $\mathbf{\Gamma}_{nn}$ is nearly singular. The direct inverse of $\mathbf{\Gamma}_{nn}$ would prove problematic and results in exceedingly large array weights. To address the problem, a simple regularization procedure is generally utilized by incorporating a positive constant to the main diagonal of the coherence matrix:[13]

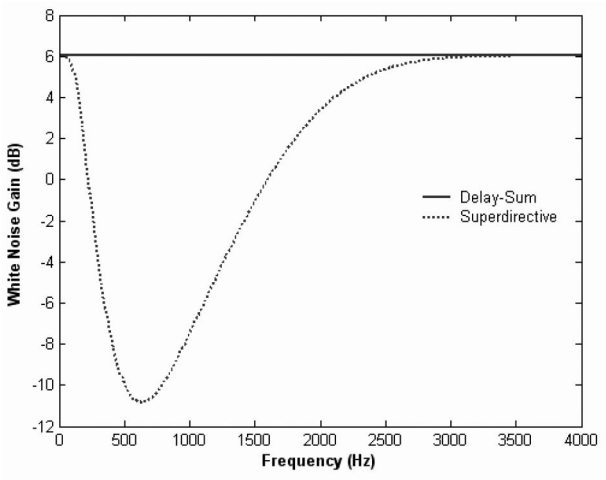
$$\mathbf{w}|_{\text{regularized}} = \frac{(\mathbf{\Gamma}_{nn} + \epsilon \mathbf{I})^{-1} \mathbf{d}}{\mathbf{d}^H (\mathbf{\Gamma}_{nn} + \epsilon \mathbf{I})^{-1} \mathbf{d}}. \tag{15}$$

The parameter ϵ can vary anywhere from zero to infinity, which corresponds to the unconstrained superdirective array or the delay-sum array, respectively. For instance, a reasonable value of ϵ to compromise the array directivity and the weight size is 0.01.

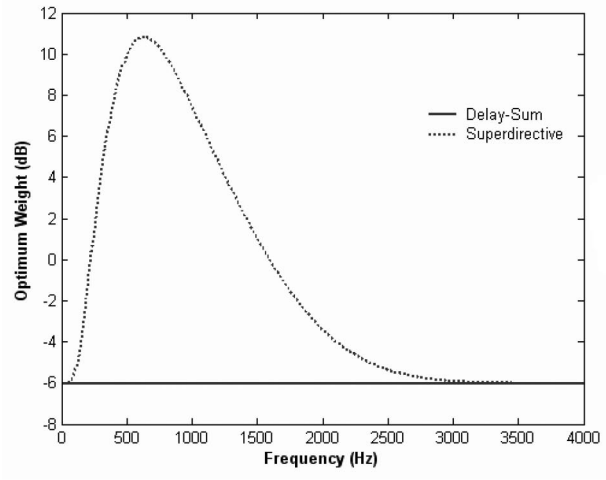
The delay-sum method and the superdirective method of a broadside array ($\theta = 0°$) are compared in terms of DI, WNG, and optimal weights as follows. Assume that there are 4 sensors equally spaced with 4 cm. Figure 2(a) illustrates the DI of the delay sum and the superdirective arrays with $\epsilon = 0.01$. Clearly visible is the improvement of DI below 3 kHz achieved using the superdirective design over the delay-sum method. Figure 2(b) compares the WNG of the delay sum and the superdirective methods with $\epsilon = 0.01$ for the broadside array. The WNG of the delay-sum array is larger than the superdirective method at the expense of low-frequency directionality. Figure 2(c) compares the 2-norm of the weight vectors of the delay sum and the superdirective methods with $\epsilon = 0.01$ for the broadside array. It can be seen from the result that the optimum weights of the superdirective array are larger than the delay-sum array at low frequencies in order to attain high directionality. The weights of the superdirective array should be maintained under a certain level in order not to create problems of filter implementation. To see more details, Fig. 3 shows the contour plots of direc-



(a)

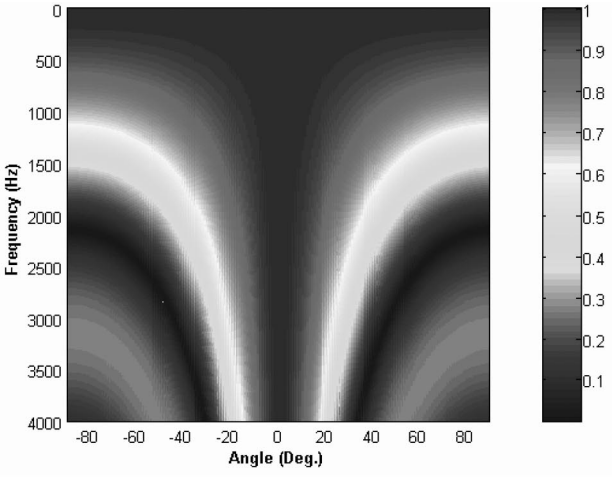


(b)

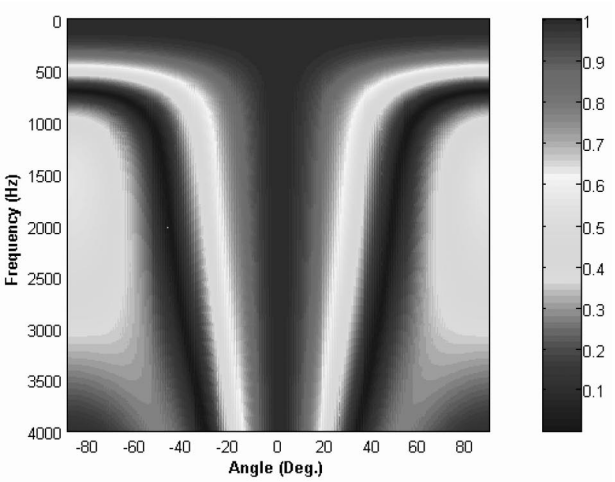


(c)

FIG. 2. A comparison of the delay-sum and the superdirective arrays with $\epsilon = 0.01$. (a) Directivity index. (b) White noise gain. (c) Optimum weight norm.



(a)



(b)

FIG. 3. The contour plots (in a linear scale) of the directional responses versus angle (x axis) and frequency (y axis) of a four-element broadside array. The interelement spacing is 4 cm. (a) Delay-sum array (b) Superdirective array.

tivity versus the angle and frequency of the delay sum and superdirective arrays. The directivity is significantly improved using the superdirective method in the band 500–1500 Hz.

The foregoing analysis based on the narrowband formulation can now be extended to the broadband scenario. The array weights obtained using the superdirective method is associated with the frequency ω. Repeating this process for the frequencies equally spaced in the band of interest gives the frequency response samples of each array filter. Now that the frequency response samples are obtained, the inverse discrete Fourier transform is applied to acquire the impulse response, or the filter coefficients, of the superdirective array. More precisely, if the P_w frequency samples of the array weight are obtained for the mth array filter, the discrete frequency response of the filter is simply

$$H_m(l) = w_m^*(l), \quad l = 1,...,P_w. \tag{16}$$

To assure real impulse responses, the frequency response samples $H_m(l)$ at $l = -1,...,-(P_w - 1)$ must be mirrored to $l = P_w + 1,...,2P_w - 1$ with proper symmetry. The FIR filter coefficients can be obtained by applying the inverse discrete Fourier transform (IDFT) to the frequency response samples for each channel:

$$h_m(k) = \frac{1}{2P_w} \sum_{l=0}^{2P_w - 1} H_m(l) W_{2P_w}^{-lk},$$

$$k = 1,...,(2P_w - 1); \quad m = 1,...,M, \tag{17}$$

where $W_{2P_w} = \exp[-j(\pi/P_w)]$. The thus obtained $h_m(k)$ is often noncausal. A circular shift with one-half the IDFT length can be performed to allow for a causal filter. Assume that there are 4 sensors equally spaced with 4 cm. The impulse responses of the superdirective array implemented using 64-tap FIR filters are shown in Fig. 4(a). The frequency responses of the superdirective filters are shown in Figs. 4(b) and 4(c). The impulse responses are symmetric and exhibit sign flipping between the microphones 1 and 2, and also the microphones 3 and 4. The frequency responses show an even more pronounced high gain and phase switching at the low-frequency range. These interesting phenomena suggest that the differential actions are necessary to produce the directivity improvement for superdirective microphones. Nevertheless, the superdirective microphones do not suffer from the poor SNR problem due to the ω^n dependence, as do pure differential microphones.[17]

Although there are many techniques available for DOA determination, we choose the simplest but most robust approach, the Fourier beamforming, due to the concern of computational loading of our processor. In the DOA session, angle spectra are calculated by steering the main beam from $-90°$ to 90° using fractional delays.[15] The maxima of angle spectra correspond to the source direction in the farfield. Figure 5 shows the angle spectra estimated at 500, 1000, 2000, and 4000 Hz by using the delay-sum and the superdirective methods, respectively. The result is plotted in a linear scale to facilitate the determination of DOA from the maxima of the angle spectra. The sound source is oriented at 10°, as indicated by the maxima of the angle spectra. The superdirective method produces a sharper DOA estimation than the delay-sum method at 500, 1000, and 2000 Hz. At 4000 Hz and higher, the superdirective method has no particular advantage. However, the side lobes of the superdirective array are larger than the delay-sum array, which is the price to pay for better resolution in DOA estimation. There is obviously a tradeoff between the beamwidth and the side lobes in the array design.

B. Performance analysis of array

Directivity is an important feature of microphone arrays, which is highly dependent on the aperture size of the arrays. In what follows, simulations and measurements were carried out to examine the directivity pattern of a 4-element linear microphone array. The signals from the microphones are summed directly without any filtering. The interelement spacing is 4 cm and the total length (aperture size) of microphone array is 12 cm. The sound source is positioned at 0°

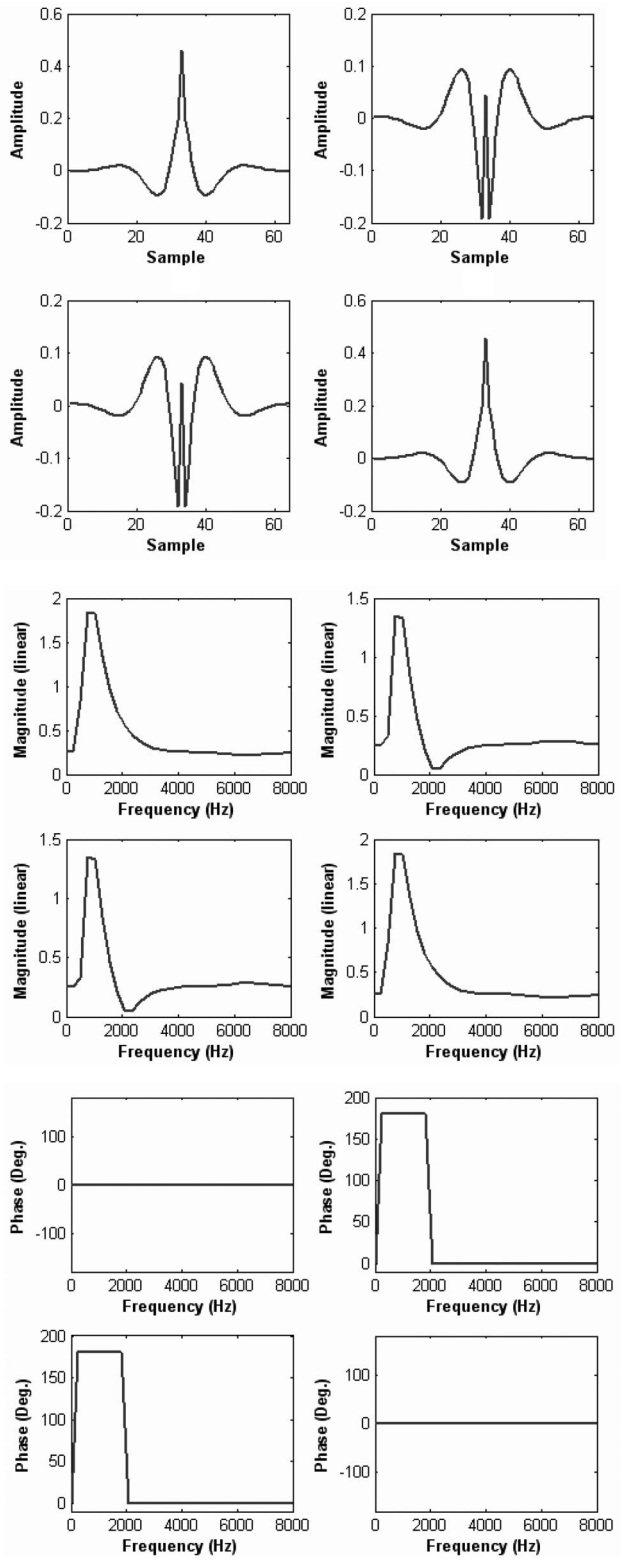


FIG. 4. The characteristics of the superdirective filter. The filter is realized as a 64-tapped FIR filter. The results are shown as 2×2 matrices, where the (1,1), (1,2), (2,1) and (2,2) correspond to the response of the 1st, 2nd, 3rd and 4th sensor, respectively. (a) The impulse response. (b) The frequency response magnitude. (c) The frequency response phase.

For reference, the simulation and measurement results of the directivity patterns of one microphone are shown in Fig. 6. The directivity pattern of one microphone at all frequencies in the simulation is omnidirectional. The directivity pattern at 500 and 1000 Hz of one microphone in the measurement are also omnidirectional, while the measurements at 2000 and 4000 Hz display slight directivity due to the effect of the baffle in which the microphones are embedded.

Figure 7 shows the simulation and measurement results of the directivity patterns of the four-element array. The directivity patterns at 500 and 1000 Hz in the simulation are nearly omnidirectional, while the directivity pattern is quite directional at 2000 and 4000 Hz. The measured data are in close agreement with the simulation. These results indicate that the direct-sum array would display directivity only above 2000 Hz. More sophisticated algorithms must be used to produce significant directivity at a low-frequency range.

Apart from directivity pattern, the SNRs of one microphone and the four-element array are also measured and listed in Table II. The SNR gain improved by a 12 cm aperture microphone array is 11.6 dB, which is close to the theoretical value of 12 dB. This suggests that the SNR can be enhanced over a single sensor by means of array structures.

III. SPATIAL SOUND RESTORATION USING HRTF

As mentioned earlier, microphone arrays have the advantage of improved SNR, directivity, and hence spatial selectivity. These features help the rejection of undesirable effects of room reverberation and acoustic feedback. However, these benefits of arrays can be countered by the loss of three-dimensional (3-D) hearing when applied to hearing aids. To address the problem, a post-filtering technique based on the Head Related Transfer Functions (HRTFs) is presented in the paper in an attempt to restore the 3-D spatial hearing mechanism resulting from the head and torso diffractions of humans. Therefore, using this microphone-HRTF system, hearing-impaired people can hear more spatial-sounding speech and music signals with a better source localization.

A. The method and system architecture

An HRTF is a measurement of the transformation for a specific source direction relative to the head, and describes the filtering process associated with the diffraction of sound by the torso, the head, and the pinnae. HRTFs contain important cues, including the interaural time differences (ITD), the interaural level differences (ILD), and spectral characteristics pertaining to spatial hearing and sound localization. ITD refers to the time difference between the left and right ears for a plane wave incident form a certain direction. ILD refers to the level difference due to the head shadowing effect between the levels of the signals received at both ears. The inverse Fourier transform of a HRTF is termed the head-related impulse responses (HRIR). A 3-D sound field can be created by convolving a source signal with the appropriate pair of HRIRs to render sound images positioned arbitrarily around a listener. The HRTF and HRIR database employed in the paper is currently available on the web.[14]

The entire system consists of a DOA estimation module, a beamsteering module, and a HRTF post-processing module. The underlying idea of the system will be explained as follows. Rewrite the array output in Eq. (2) with a discrete-time and broadband setting,

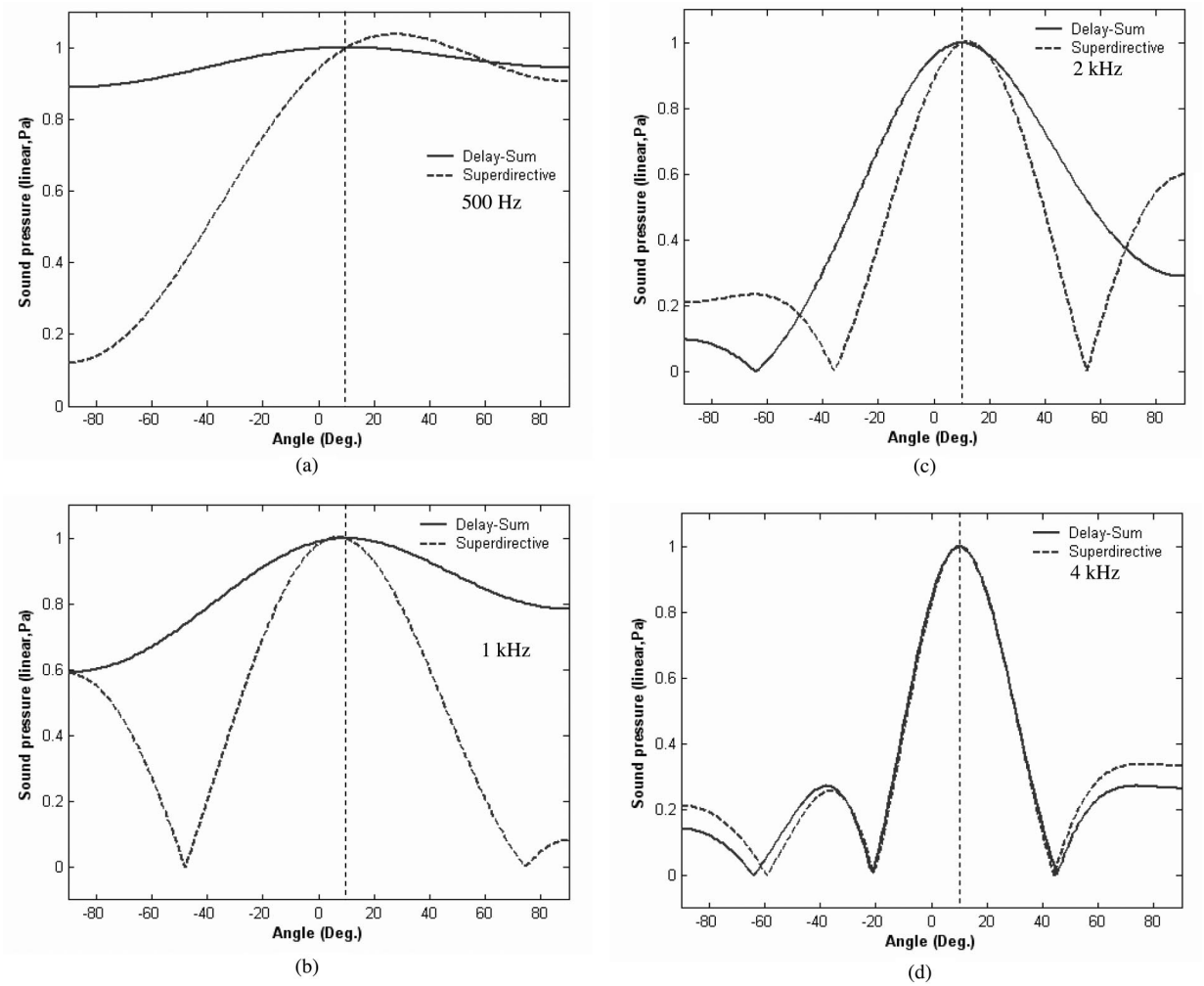


FIG. 5. The simulation results of the DOA estimation (in a linear scale) by the delay-sum and superdirective methods. The sound source is a pure tone and is located at $\theta=10°$, as indicated by a vertical dotted line in the figure. (a) 500 Hz; (b) 1000 Hz; (c) 2000 Hz; (d) 4000 Hz.

$$y(n)=\mathbf{w}^H(n)*\mathbf{x}(n), \tag{18}$$

where* denotes convolution and n is the discrete-time index. In Eq. (17),

$$\mathbf{x}(n)=s(n)*\mathbf{a}(\theta,n) \tag{19}$$

being the data vector received at the microphones, and $\mathbf{a}(\theta,n)$ is the impulse response of the array manifold vector associated with the source direction θ. After a DOA session using the superdirective beamformer, a potential source direction is found. Then, the beam of the array is electronically steered to that direction using the second-order Lagrange interpolation for the steering vector $\mathbf{d}\approx\mathbf{a}$. Note that the beam-steering process should guarantee $\mathbf{w}^H\mathbf{d}\approx 1$. Although this is a fixed beamformer, we choose the simple but practical approach for two reasons. First, apart from processing needs in multichannel array filtering, DOA, HRTF, and real-time display, there is really not much computational power left in the present DSP system to afford sophisticated adaptive algorithms such as the generalized sidelobe canceller (GSC) or the Griffiths beamformer.[7] Second, adaptive beamformers still have robustness issues in the context of a steering vector error and correlated noise.[13] In particular, the latter issue frequently occurs in a live room environment such as a car cabinet, where multipath reflections could lead to an undesirable signal-noise cancellation.

In order to recover the spatial sound image for the array system, as a final step, HRIRs are used to filter the array output to yield the binaural signals for a headphone,

$$\mathbf{y}(n)=\begin{bmatrix} h_L(n) \\ h_R(n) \end{bmatrix}^* y(n), \tag{20}$$

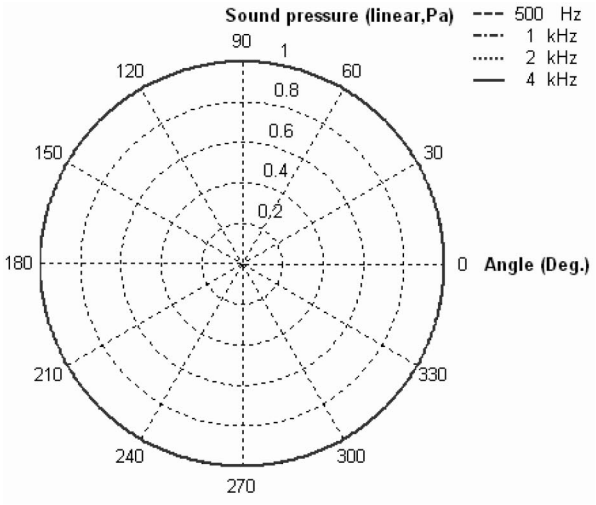
where $h_L(n)$ and $h_R(n)$ are the HRIRs of the left and the right ears. In effect, the overall output that the system will produce is

$$\mathbf{y}(n)=\begin{bmatrix} h_L(n) \\ h_R(n) \end{bmatrix}^* s(n)*\mathbf{w}^H(n)*\mathbf{a}(\theta,n). \tag{21}$$

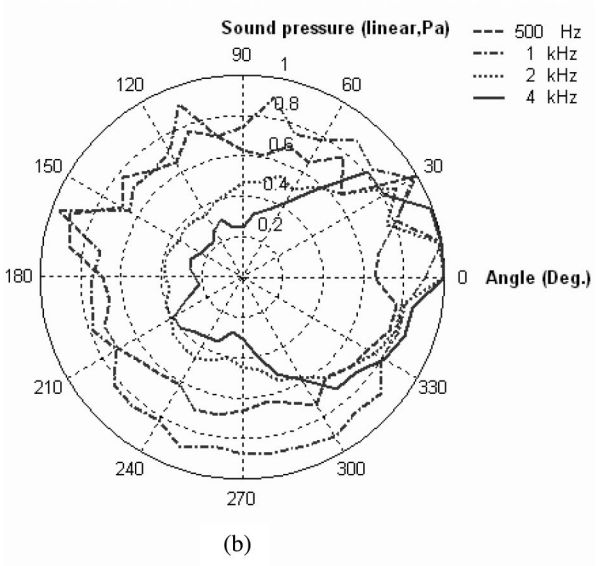
The preceding procedure of post-processing for 3-D hearing is summarized in the flow chart of Fig. 8.

B. Experimental verifications

The above-mentioned array signal processing algorithms for 3-D spatial sound restoration with HRTF was implemented on a digital signal processor (DSP), TMS320C32. Figure 9 shows the system block diagram, including a computer equipped with the DSP, a sound source, a four-element

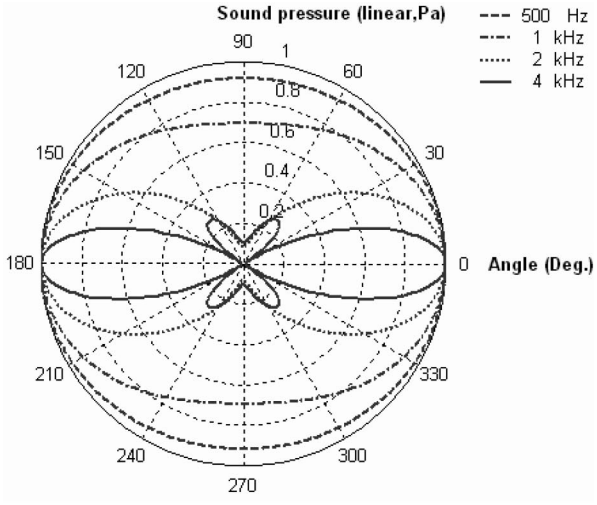


(a)

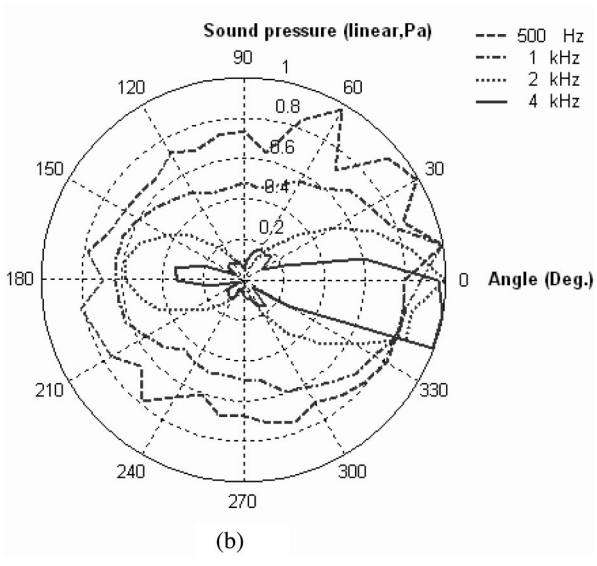


(b)

FIG. 6. The directivity pattern of one microphone. (a) Simulation. (b) Measurement.



(a)



(b)

FIG. 7. The directivity pattern of a broadside delay-sum array with four microphones equally spaced by 4 cm. (a) Simulation. (b) Measurement.

linear microphone array, a preamplifier, and two second-order antialiasing filters with cutoff frequency 8 kHz and a headphone. Four condenser microphones with 6 mm diameter are fitted in a acrylic plastic plate and the interelement spacing of microphones is 4 cm, which gives a total length of 12 cm. The sound source is placed at $\theta=10°$. The sampling rate was chosen to be 16 kHz.

The delay-sum beamformer and the superdirective beamformer are applied to estimate DOA using the DSP system. Figure 10 shows the experimental results of the DOA estimation of obtained using the delay sum and the superdirective methods at 500, 1000, 2000, and 4000 Hz, respectively. A more accurate DOA estimation at 500, 1000, and 2000 Hz can be obtained by using the superdirective method than the delay-sum method because of the narrower beamwidth of the former. This observation is in agreement with the simulation results. At 4000 Hz, two beamformers yielded a comparable performance. Figure 11 shows the DOA estimation obtained using the delay-sum and the superdirective methods, in which case a sound source located at $\theta=10°$ is a random noise bandlimited to 8 kHz. An error analysis of DOA estimation obtained experimentally by the delay-sum beamformer and the superdirective beamformer is illustrated in Table III. The average DOA error of the superdirective beamformer is 0° vs 6° obtained using the delay-sum beamformer. The former method significantly outperforms the latter one, owing to its narrower beamwidth (angle resolution).

After the DOA was estimated, the beamformer in conjunction with beam steering was utilized to enhance the sound signals. Only the superdirective beamformer was applied in the following tests because of its superior angle resolution. The fractional delays should be carefully compensated using Lagrange interpolation. Otherwise, a large

TABLE II. The measured SNRs and SNR gain of the single microphone versus the four-element microphone array with with a 12 cm aperture.

Aperture size	Microphone number	SNR (dB)	SNR gain (dB)
12 cm	One	46.1	11.6
	Four	57.7	

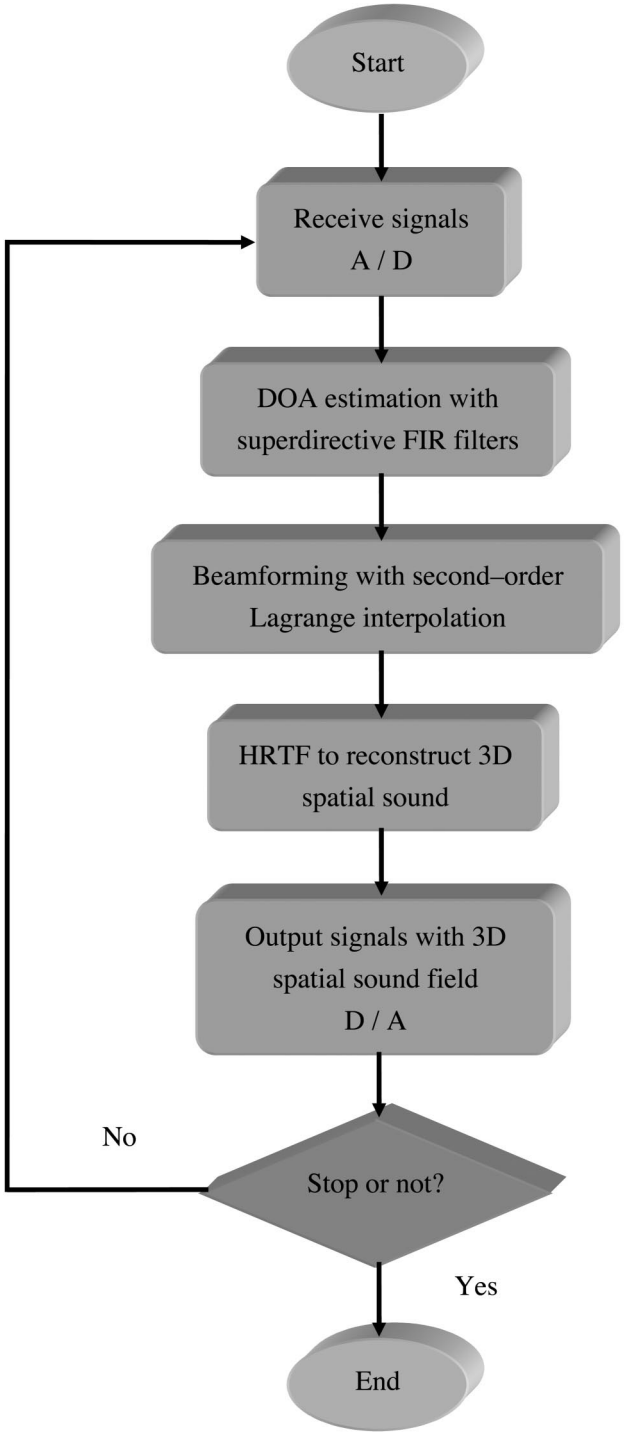


FIG. 8. The flow chart of array signal processing with 3-D spatial sound restoration with the HRTF.

steering error would arise, especially when the sampling rate is low.

Following the beam-steering session, post-processing based on HRTFs was carried out to realize the 3-D spatial sound. When the location of the sound source is changed, HRTFs were also updated according to the new direction found in the DOA session. In order to evaluate the effectiveness of the system in localizing sound sources, a subjective experiment was conducted. The listening test involved ten human subjects. Random noise bandlimited to 8 kHz was

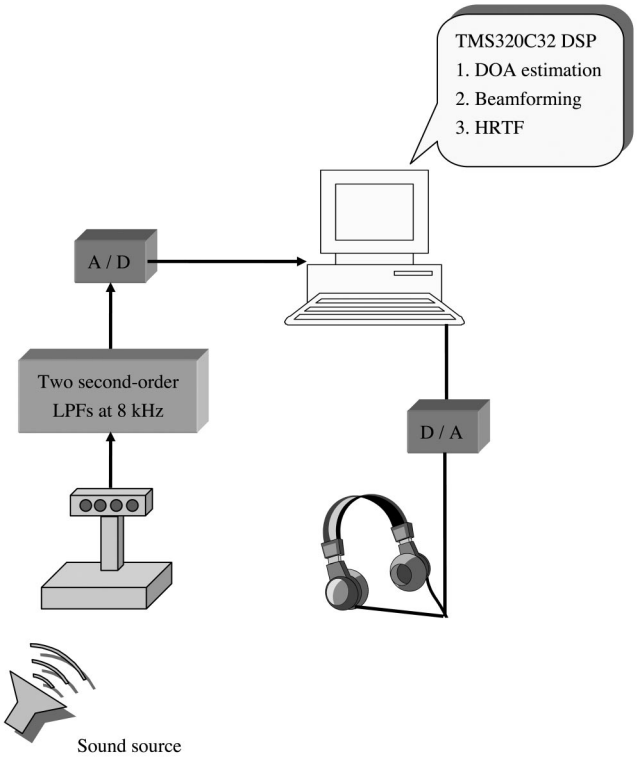


FIG. 9. The block diagram of the array spatial sound system implemented on a DSP, TMS320C32.

used as the source. A headphone was employed as the rendering device. In the experiment, 11 positions, ranging from $-50°$ to $50°$ with $10°$ intervals, were preselected to position the source. The distance between the microphone array and the source was 2 m. The experiment was conducted in an anechoic chamber to minimize unwanted reflections. Figure 12 shows the result of the subjective localization experiment. The perceived angles of source are in very good agreement with the presented angles of the source since most data points fall on the diagonal of the plot. Figure 13 shows the statistics of localization errors in a bar chart. It can be seen in the result that the localization errors at $\pm 10°$ and $\pm 50°$ are somewhat larger than the other angles. Overall, the average localization error is 11.7° (nearly one test interval), and the standard deviation is 6.5°. It is noted that the DOA estimation is quite accurate using the superdirective beamformer (0° from Table III), and the discrepancy of the subjective test is due predominantly to the HRTF database. In summary, these results reveal that the developed system is effective in creating a spatial sound field that allows for the practical localization needs of human listeners.

IV. CONCLUSIONS

A microphone array accompanied with 3-D spatial post-processing system has been developed in this paper. Array signal processing algorithms are implemented on a DSP system to accommodate broadband acoustical applications with better SNR and directivity. To further enhance the directivity at the low-frequency range, the superdirective method is exploited in the beamformer design. The entire system consists of a DOA estimation module, a beam steering module, and

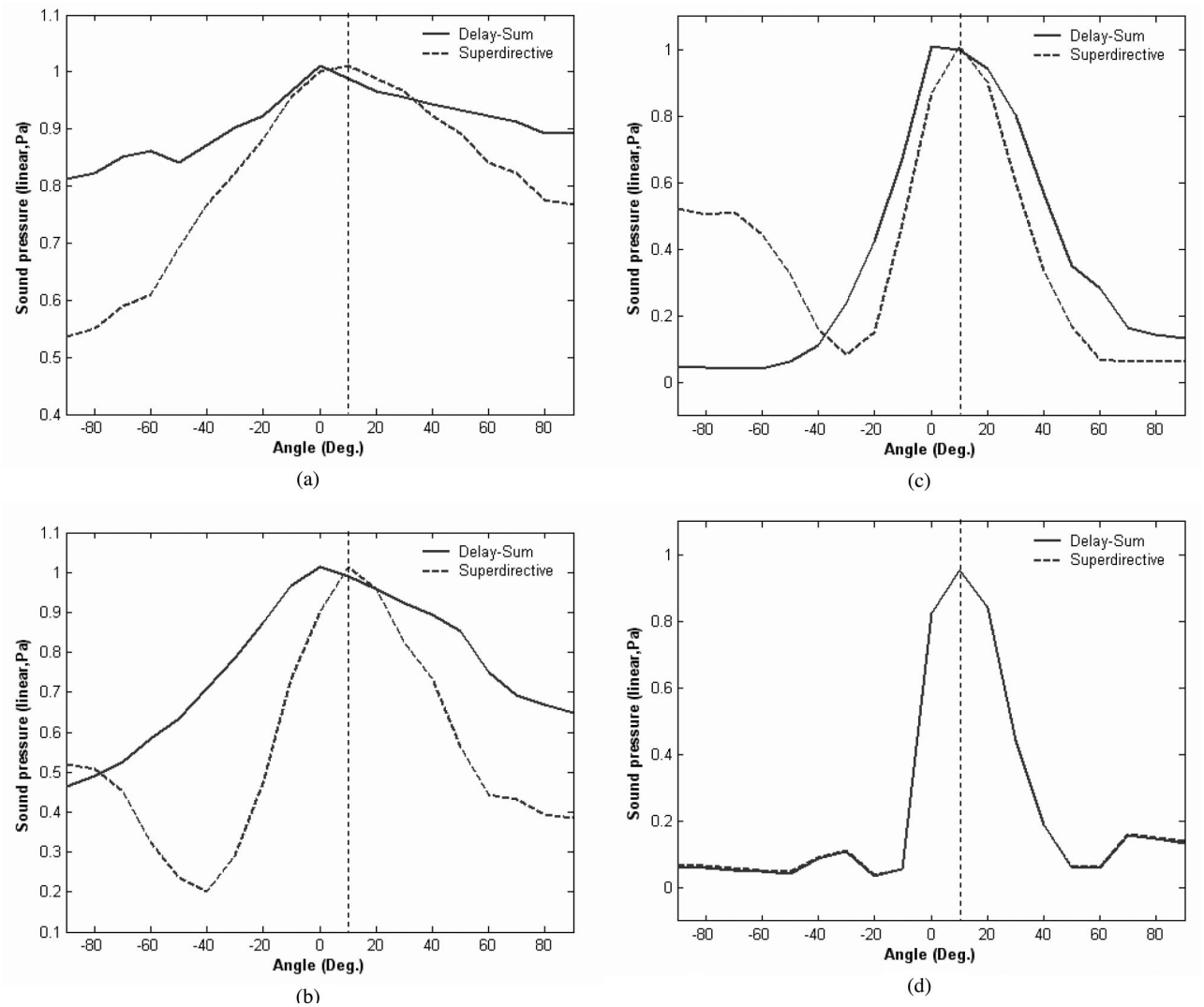


FIG. 10. The experimental results of the DOA estimation (in a linear scale) by the delay-sum and superdirective methods. The sound source is a pure tone and is located at $\theta=10°$, as indicated by a vertical dotted line in the figure. (a) 500 Hz. (b) 1000 Hz. (a) 2000 Hz. (b) 4000 Hz.

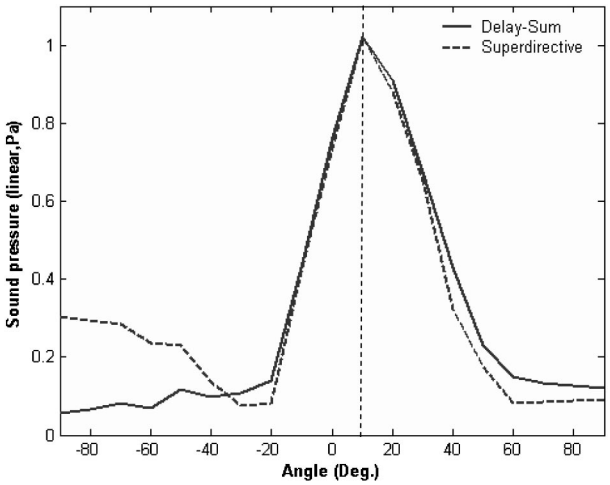


FIG. 11. The experimental results of the DOA estimation by the delay-sum and superdirective methods. The sound source is a random noise bandlimited to 8 kHz and located at $\theta=10°$, as indicated by a vertical dotted line in the figure.

an HRTF post-processing module. It was suggested by the reviewer that, instead of free-field HRTFs, stero room responses should be used. However, the present paper did not choose this approach because we feel that the natural reverberations and responses of the room environment were not lost in beamforming. What had been lost was the information associated with human head scattering and diffraction. In addition, as another practical reason, filtering of room responses is known to be a computationally expensive operation, which is still prohibitive in the present DSP platform used in the research.

TABLE III. Error analysis of DOA estimation obtained by the delay-sum beamformer and the superdirective beamformer.

Source	Delay-sum beamformer	Superdirective beamformer
500 Hz sine	−10°	0°
1 kHz sine	−10°	0°
2 kHz sine	−10°	0°
4 kHz sine	0°	0°
Broadband random	0°	0°
Average error	6°	0°

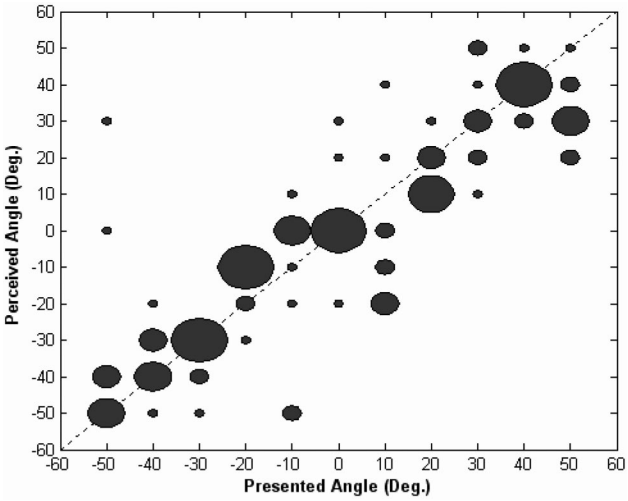


FIG. 12. The results of the subjective listening experiment obtained using the proposed array spatial sound system. The perceived angle (y axis) is plotted versus the presented angle (x axis). The 45° line stands for perfect localization.

Simulations and experiments were carried out to investigate the feasibility of the proposed array system. It was found from the results that the array has achieved significant improvement in terms of SNR as well as directivity. A subjective listening experiment was conducted to examine the source localization performance using the present system. The results indicate that the subjects are capable of localizing the presented source direction within an average error of 11.7°.

Although these preliminary tests reveal the potential of the present system, several limitations of the work should be mentioned. First, the selected sampling rate of 16 kHz was somewhat low because of the limited processing power of the present DSP we used. This gives an effective frequency range of less than 8 kHz, which might be insufficient for sound signals other than speech. However, this limitation is minor and can be removed by using a faster DSP. Second, the system works quite well with continuous and stationary signals but may lose track of the source for transient signals. Tracking algorithms suited to transient signals remain to be investigated. Third, background noise and reflection in a live room may seriously interfere the DOA estimation of the present system—a problem common to array systems. It is worth exploring in the future research algorithms for the DOA estimation that are both sharp and robust in a reverberant environment.

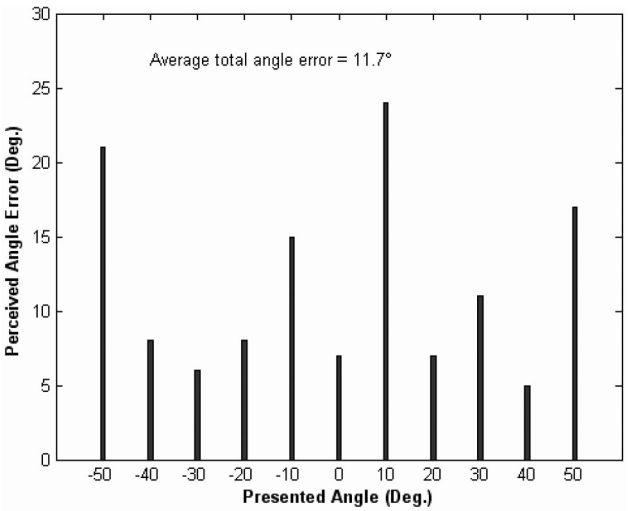


FIG. 13. The bar chart showing the statistics of localization errors of the subjective listening experiment for each presented angle. The average localization error is 11.7° and the standard deviation is 6.5°.

ACKNOWLEDGMENT

The work was supported by the National Science Council in Taiwan, Republic of China, under Project No. NSC 92-2212-E009-030.

[1] P. L. Chu, "Superdirective microphone array for a set-top videoconference system," *Proceedings of the IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 1997.
[2] D. Korompis, K. Yao, R. E. Hudson, and F. Lorenzelli, "Microphone array signal processing for hearing aid application," *IEEE Workshop on VLSI SP*, Chicago, 1994.
[3] J. M. Kates and M. R. Weiss, "A comparison of hearing-aid array processing techniques," J. Acoust. Soc. Am. **99**, 3138–3148 (1996).
[4] D. Korompis, A. Wang, and K. Yao, "Comparison of microphone array designs for hearing aid," *Proceedings of the IEEE ICASSP*, 1995.
[5] C. Liu and S. Sideman, "Simulation of fixed microphone arrays for directional hearing aids," J. Acoust. Soc. Am. **100**, 848–856 (1996).
[6] G. H. Saunders and J. M. Kates, "Speech intelligibility enhancement using hearing-aid array processing," J. Acoust. Soc. Am. **102**, 1827–1837 (1997).
[7] H. L. Van Trees, *Optimum Array Processing* (Wiley, New York, 2002).
[8] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **35**, 1365–1375 (1987).
[9] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," IEEE Trans. Acoust., Speech, Signal Process. **34**, 393–398 (1986).
[10] J. M. Kates, "Superdirective arrays for hearing aids," J. Acoust. Soc. Am. **94**, 1930–1943 (1993).
[11] R. W. Stadler and W. Rabinowitz, "On the potential of fixed arrays for hearings aids," J. Acoust. Soc. Am. **94**, 1332–1342 (1993).
[12] J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "An alternative implementation of the superdirective beamformer," *Proceedings of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, 1999, pp. 17–20.
[13] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications* (Springer-Verlag, New York, 2001).
[14] B. Gardner and K. Martin, "HRTF measurements of a KEMER dummy-head microphone," MIT Media Lab., 1994.
[15] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay," *IEEE Signal Processing Magazine*, January, 1996, pp. 30–60.
[16] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**, 926–935 (1972).
[17] G. W. Elko, "Superdirectional microphone arrays," *Acoustical Signal Processing for Telecommunication*, edited by S. Gay and J. Benesty (Kluwer Academic, Boston, 2000).