Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction

Mingsian R. Bai[a] and Chih-Chung Lee
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

(Received 2 December 2005; revised 27 June 2006; accepted 29 June 2006)

A comprehensive study was conducted to explore the effects of listening angle on crosstalk cancellation in spatial sound reproduction using two-channel stereo systems. The intention is to establish a sustainable configuration of crosstalk cancellation system (CCS) that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice. Although crosstalk can in principle be suppressed using multichannel inverse filters, the CCS does not lend itself very well to practical application owing to the fact that the sweet spot is being so small. Among the parameters of loudspeaker deployment, span angle is a crucial factor that has a profound impact on the separation performance and sweet spot robustness achievable by the CCS. This paper seeks to pinpoint, from a more comprehensive perspective, the optimal listening angle that best reconciles the robustness and performance of the CCS. Two kinds of definitions of sweet spot are employed for assessment of robustness. In addition to the point source model, head related transfer functions (HRTF) are employed as the plant models in the simulation to emulate more practical localization scenarios such as the high-frequency head shadowing effect. Three span angles including 10, 60, and 120 deg are then compared via objective and subjective experiments. The Friedman test is applied to analyze the data of subjective experiments. The results indicate that not only the CCS performance but also the panning effect and head shadowing will dictate the overall performance and robustness. The 120-deg arrangement performs comparably well as the standard 60-deg arrangement, but is much better than the 10-deg arrangement. [DOI: 10.1121/1.2257986]

PACS number(s): 43.38.Md, 43.60.Tj, 43.60.Pt [AJZ] Pages: 1976–1989

I. INTRODUCTION

The central idea of spatial audio reproduction is to synthesize a virtual sound image. The listener perceives as if the signals reproduced at the listener's ears would have been produced by a specific source located at an intended position.[1,2] This attractive feature of spatial audio lends itself to an emerging audio technology with promising application in mobile phone, personal computer multimedia, video games, home theater, etc.

The rendering of spatial audio is either by headphones or loudspeakers. Headphones reproduction is straightforward, but suffers from several shortcomings such as in-head localization, front-back reversal, and discomfort to wear. While loudspeakers do not have the same problems as the headphones, another issue adversely affects the performance of spatial audio rendering using loudspeakers. The issue frequently encountered in loudspeaker reproduction is the crosstalk in the contralateral paths from the loudspeakers to the listener's ears, which may obscure source localization. To overcome the problem, crosstalk cancellation systems (CCS) that seek to minimize, if not totally eliminate, the crosstalks have been studied extensively by researchers.[3–8] Various inverse filtering approaches were suggested for designing multichannel prefilters for CCS.

Notwithstanding the preliminary success of CCS in an academic community, a problem seriously hampers the use of CCS in practical applications. The problem stems from the limited size of the so-called "sweet spot" in which CCS remains effective. The sweet spots are generally so small especially at the lateral side that a head movement of a few centimeters would completely destroy the cancellation performance. Two kinds of approaches can be used to address this problem—the adaptive design and the robust design. An example of adaptive CCS with head-tracker was presented in the work of Kyriakakis *et al.*[9,10] This approach dynamically adjusts the CCS filters by tracking the head position of the listener using optical or acoustical sensors. However, the approach has not been widely used because of the increased hardware and software complexity of the head tracker. On the other hand, instead of dynamically tracking the listener's head, an alternative CCS design using fixed filters can be taken to create a "widen" sweet spot that accommodates larger head movement. Ward and Elko in Bell Labs have conducted a series of insightful analysis of the robustness issue of CCS. In their paper[11] on this topic in 1998, robustness of a two-channel stereo loudspeaker (2×2) CCS was investigated using weighted cancellation performance measure at the pass zone and stop zone, respectively. In the other paper[12] by the same authors in 1999, robustness issue of a 2×2 CCS was revisited using a different measure, the condition number, which focuses more on numerical stability during matrix inversion, in the presence of noise in data

[a]Elecrtonic mail: msbai@mail.nctu.edu.tw

and/or perturbations to system properties. Yet, in another paper[13] by Ward, a joint least squares optimization method is employed to obtain a CCS that is robust to head misalignment. The above-mentioned research winds up with a simple but important conclusion that the optimal loudspeaker spacing should be inversely proportional to the operating frequency. Along the line of robust CCS design, a celebrated "stereo dipole", configuration was suggested by Kirkeby, Nelson, and Hamada[14] and Takeuchi and Nelson.[15] In their arrangement, two loudspeakers are closely spaced with only a 10° span. Their analysis of robustness of CCS also focused primarily on numerical stability in relation to the errors in matrix inversion. The consistent finding of these studies was that the optimal loudspeaker spacing is inversely proportional to the operating frequency. Since the optimal spacing is frequency dependent, a multidrive configuration of the optimal source distribution (OSD) system,[16] comprising pairs of loudspeakers with various spacings, was suggested to deal with crosstalks for different frequency bands. Another multidrive CCS design was also developed by Bai *et al.*[18] based on the genetic algorithm and array signal processing. Their approach requires no crossover circuits as in the OSD system.

According to Gardner,[19] loudspeakers spaced apart tend to yield a smaller equalization zone than loudspeakers spaced closely. However, the improvement is predominantly along the front-back axis and the equalization zone widens only slightly when the speakers are positioned closely together. One disadvantage of close spacing is the lack of natural high frequency separation due to head shadowing. Another problem is that small head rotation will cause both speakers to fall on the same side so that the panning mechanism fails.

Thus far, there have been pros and cons in the closely spaced CCS. The question of which kind of loudspeaker arrangement is the best has been puzzling people for quite some time. It is worth exploring further the underlying physical insights from all possible angles. This motivates the current research to undertake a comprehensive study in a hope to resolve this optimal CCS problem more conclusively. In Gardner's work,[19] the head-related transfer functions (HRTF) were measured in the MIT Media Lab[20,21] and subjective listening tests were conducted. However, only the crosstalk below 6 kHz was considered to result in a band-limited CCS design. Furthermore, the robustness of CCS to head misalignment were discussed in depth by Takeuchi and Nelson.[15] In both works, only two listening spans including 10- and 60-deg spans were investigated. On the other hand, the emphasis of this paper is placed on the analysis of the effects of listening angle on CCS in terms of not only robustness but also performance. There are several special features in this paper. First, not only the robustness but also the performance of CCS is examined with the aid of a more comprehensive set of indices. Second, two kinds of definitions of sweet spot are employed for assessment of robustness. Third, the present work considers the entire audible 20 kHz band in which the listener's head may provide natural separation for certain loudspeaker arrangements. Fourth, apart from the objective physical tests, subjective listening tests are conducted to practically assess the CCS arrangements with different listening angles. The results of subjective tests will be validated by using the Friedman test. Although the last three points have been investigated in Refs. 15 and 19, this study examines the design issues in further detail and in some cases reaches different conclusions than the previous research. The intention is to establish a sustainable configuration of CCS that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice.

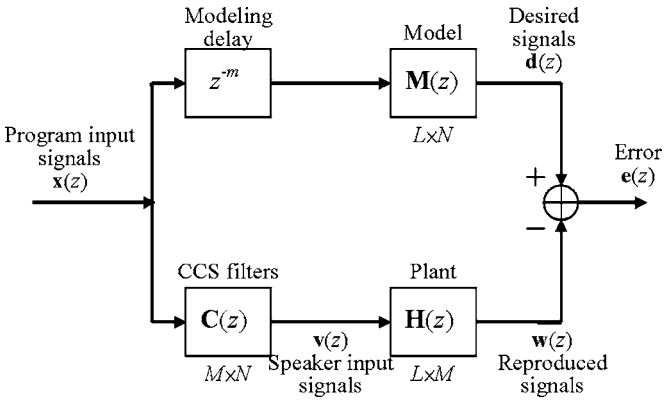

FIG. 1. The block diagram of a multichannel model-matching problem in the CCS design.

II. MULTICHANNEL INVERSE FILTERING FOR CCS FROM A MODEL-MATCHING PERSPECTIVE

The CCS aims to cancel the crosstalk in the contralateral paths from the multichannel loudspeakers to the listener's ears so that the binaural signals are reproduced at two ears like those reproduced using headphones. This problem can be viewed from a model-matching perspective, as shown in Fig. 1. In the block diagram, $\mathbf{x}(z)$ is a vector of N program input signals, $\mathbf{v}(z)$ is a vector of M loudspeaker input signals, and $\mathbf{e}(z)$ is a vector of L error signals. $\mathbf{M}(z)$ is a $L\times N$ matrix of matching model, $\mathbf{H}(z)$ is a $L\times M$ plant transfer matrix, and $\mathbf{C}(z)$ is a $M\times N$ matrix of the CCS filters. The z^{-m} term accounts for the modeling delay to ensure causality of the CCS filters. Let us neglect the modeling delay for the moment, it is straightforward to write down the input-output relationship

$$\mathbf{e}(z)=[\mathbf{M}(z)-\mathbf{H}(z)\mathbf{C}(z)]\mathbf{u}(z). \tag{1}$$

For arbitrary inputs, minimization of the error output is tantamount to the following optimization problem:

$$\min_{\mathbf{C}}\|\mathbf{M}-\mathbf{HC}\|_F^2, \tag{2}$$

where F symbolizes the Frobenius norm.[22] For a $L\times N$ matrix $\mathbf{A}$, Frobenius norm is defined as

$$\|\mathbf{A}\|_F^2=\sum_{n=1}^{N}\sum_{l=1}^{L}|a_{ln}|^2$$

$$=\sum_{n=1}^{N}\|\mathbf{a}_n\|_2^2,\ \mathbf{a}_n \text{ begin the } n\text{th column of } \mathbf{A}. \tag{3}$$

Hence, the minimization problem of Frobenius norm can be

converted to the minimization problem of two norm by partitioning the matrices into columns. Assume that **H** is of full column rank and there is no coupling between the columns of the resulting matrix **C** which approximates the inverse of **H**, the minimization of the square of the Frobenius norm of the entire matrix **H** is tantamount to minimizing the square of each column independently. Therefore, Eq. (2) can be equal to the following equation:

$$\min_{\mathbf{c}_n, n=1,2,\ldots,N} \sum_{n=1}^{N} \|\mathbf{H}\mathbf{c}_n - \mathbf{m}_n\|_2^2, \tag{4}$$

where $\mathbf{c}_n$ and $\mathbf{m}_n$ are the nth column of the matrices **C** and **M**, respectively. The optimal solution of $\mathbf{c}_n$ can be obtained by applying the method of least squares to each column

$$\mathbf{c}_n = \mathbf{H}^+\mathbf{m}_n, \quad n = 1,2,\ldots,N, \tag{5}$$

where $\mathbf{H}^+$ is the pseudoinverse of **H**.[22] This optimal solution in the least-squares sense can be assembled a more compact matrix form

$$[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad \mathbf{c}_N] = \mathbf{H}^+[\mathbf{m}_1 \quad \mathbf{m}_2 \quad \cdots \quad \mathbf{m}_N] \tag{6a}$$

or

$$\mathbf{C} = \mathbf{H}^+\mathbf{M}. \tag{6b}$$

For a matrix **H** with full-column rank ($L \geq M$), $\mathbf{H}^+$ can be calculated according to

$$\mathbf{H}^+ = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H. \tag{7}$$

Here, $\mathbf{H}^+$ is also referred to as the left pseudoinverse of **H** in that $\mathbf{H}^+\mathbf{H} = \mathbf{I}$.

In practice, the number of loudspeakers is usually greater than the number of ears, i.e., $L \leq M$. Regularization can be used to prevent the singularity of $\mathbf{H}^H\mathbf{H}$ from saturating the filter gains.[23,24]

$$\mathbf{H}^+ = (\mathbf{H}^H\mathbf{H} + \beta\mathbf{I})^{-1}\mathbf{H}^H. \tag{8}$$

The regularization parameter β can either be constant or frequency dependent.[25] It is noted that the procedure to obtain the filter **C** in Eq. (6) is essentially a frequency-domain formulation, inverse Fourier transform along with circular shift (hence the modeling delay) are needed to obtain causal FIR filters.

III. NUMERICAL SIMULATIONS

In this section, numerical simulations are conducted to examine the effects that listening angle has on CCS. The free-field point source model and HRTFs are employed as the plant models in the simulations. Only lateral misalignment is considered because it has been concluded by the previous research that the lateral misalignment has more pronounced effect on CCS than the other types of head movements.[15]

A. Free-field point source model

For the free-field point source model illustrated in Fig. 2, the plant transfer matrix can be shown to be

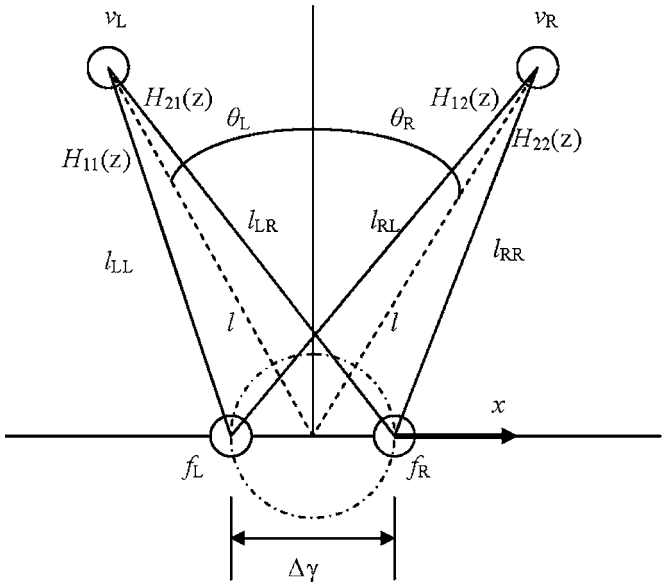

FIG. 2. The geometry of the free-field point source model.

$$\mathbf{H} = \frac{\rho_0}{4\pi}\begin{bmatrix} e^{-jk_a l_{LL}/l_{LL}} & e^{jk_a l_{RL}/l_{RL}} \\ e^{-jk_a l_{LR}/l_{LR}} & e^{-jk_a l_{RR}/l_{RR}} \end{bmatrix} k_a = \omega/c_0, \tag{9}$$

where k_a, ρ_0, and c_0 represent the wave number, the density, and sound speed, respectively. In the simulation, we assume that $c_0 = 343$ m/s, $\rho_0 = 1.21-$ kg/m^3, $l = 1.4$ m, and the spacing between ears. $\Delta\gamma = 0.1449$ m.[26] In Eq. (9), the lengths are calculated as

$$l_{LL} = \left[(l\cos\theta)^2 + \left(l\sin\theta - \frac{\Delta\gamma}{2} + x \right)^2 \right]^{1/2}, \tag{10a}$$

$$l_{LR} = \left[(l\cos\theta)^2 + \left(l\sin\theta + \frac{\Delta\gamma}{2} + x \right)^2 \right]^{1/2}, \tag{10b}$$

$$l_{RL} = \left[(l\cos\theta)^2 + \left(l\sin\theta + \frac{\Delta\gamma}{2} - x \right)^2 \right]^{1/2}, \tag{10c}$$

$$l_{RR} = \left[(l\cos\theta)^2 + \left(l\sin\theta - \frac{\Delta\gamma}{2} - x \right)^2 \right]^{1/2}. \tag{10d}$$

The CCS filters are obtained by using the aforementioned inverse filtering procedure with constant regularization parameters. Overall, 256 frequencies equally spaced from 20 to 20 kHz on a logarithmic frequency scale are selected. The kth selected frequency can be represented as

$$f(k) = 10^{\log_{10}^{20} + (\log_{10}^{20\,000} - \log_{10}^{20})k/256}, \quad k = 0,1,\ldots,255, \tag{11}$$

where $\log_{10}^{20}$ and $\log_{10}^{20\,000}$ symbolize the logarithm with base 10 for 20 Hz and 20 kHz, respectively. In the simulation, the power of each CCS filter at different span angles is constrained to be equal, which can be achieved by using different regularization values. The 2×2 transfer function matrix is assumed to be symmetric. The power of CCS filters is defined as

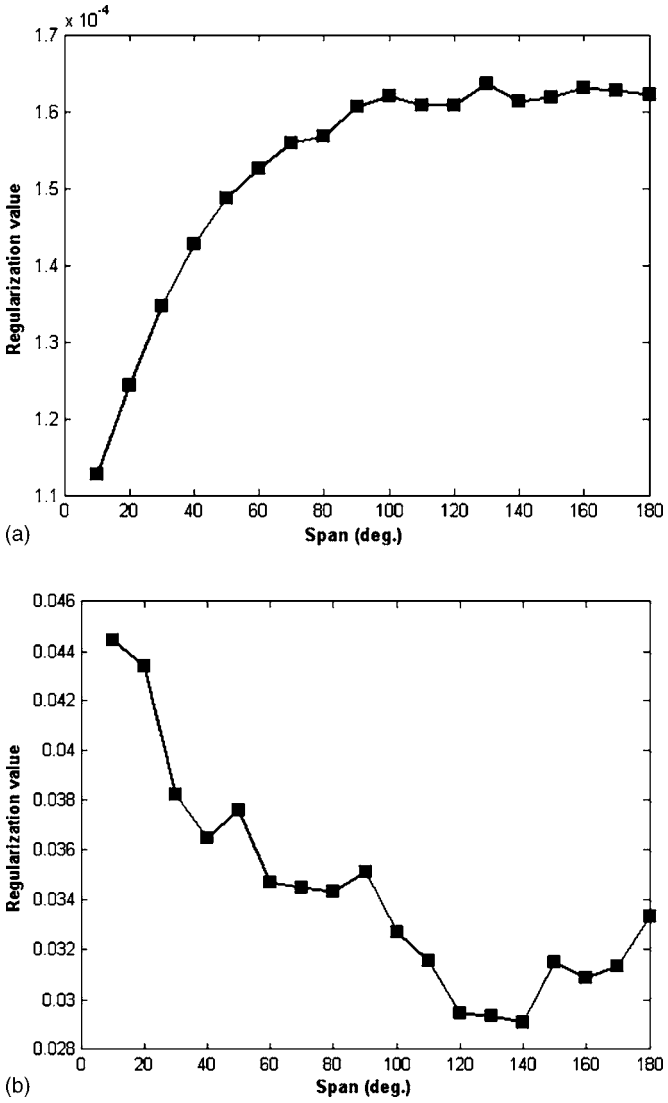

FIG. 3. The values of regularization in (a) the free-field point source model and (b) the HRTF model.

$$\frac{1}{P}\sum_{k=0}^{P-1}[|C_{11}(k)|^2+|C_{12}(k)|^2], \tag{12}$$

where C_{11} and C_{12} are diagonal and off-diagonal component of the CCS filter, P is the number of frequency samples and k represents the frequency index. The regularization values in each span angle are shown in Fig. 3(a).

Let the overall response of the CCS filters cascaded with the acoustic plant be

$$\mathbf{G}=\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}=\mathbf{HC}. \tag{13a}$$

Channel separation, defined as the ratio of the contralateral response and the ipsilateral response compensated by CCS, is employed as a performance index

$$CHSP_{\mathrm{L}}(k)=G_{12}(k)/G_{11}(k) \quad \text{or } CHSP_{\mathrm{R}}(k) = G_{21}(k)/G_{22}(k). \tag{13b}$$

Figure 4(a) shows the contour plot of the condition number of the plant matrix **H** in the nominal center position ($x=0$). The x axis is the listening angles in degrees and the y axis is logarithmic frequency in hertz. Condition number in decibels is represented by gray levels. In addition, the contour plots of the filter gain and the channel separation shown in Figs. 4(b) and 4(c) are plotted versus the same coordinates as in Fig.

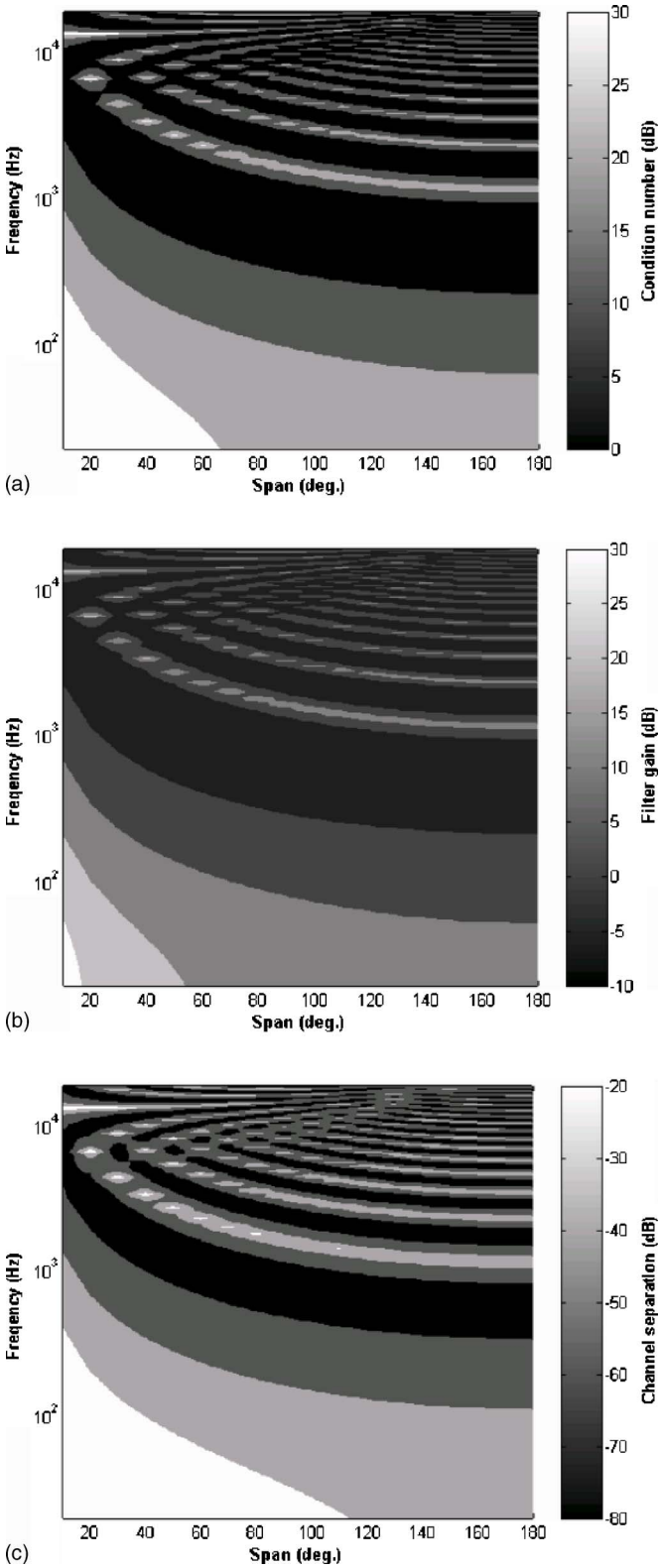

FIG. 4. The contour plots calculated using the point source model of (a) the condition number of acoustical plant matrix **H**, (b) the filter gain, and (c) the channel separation.

4(a). From the plots, the condition number follows a similar trend to the filter gain. This reveals that there is indeed a tradeoff between numerical stability and separation performance. Specifically, a large condition number leads to high filter gain. This in turn calls for regularization to restrain the filter gain at the compromise of some performance.

Another issue of CCS is concerned with the *ringing frequency* given by[15–17]

$$f_n = \frac{nc}{2\Delta r \sin\theta}, \quad n = 0,1,2,\ldots,. \tag{14}$$

Ringing frequencies appear at high frequency particularly for small span arrangement. Suppose the frequency range of our interest is from 100 to 6 kHz. Although the 10-deg span arrangement is well conditioned at frequencies below the intersection of the 6 kHz line and the first ringing, it suffers from the "corner problem," where poor conditioning and high gain arise at low frequencies and small spans. This is to be expected because the acoustic plants are almost identical in magnitude and phase when the listening angle becomes exceedingly small.

Figures 5(a)–5(c) show the contour plots of channel separation at the right ear for three span angles (2θ), 10, 60, and 120 deg, respectively. The span of 10 and 60 deg are selected because they correspond to stereo dipole and International Telecommunications Union (ITU) standard.[27] The x axis is the lateral head displacements in centimeters and the y axis is logarithmic frequency in hertz. Channel separation in decibel is represented by gray levels. The darker the gray level, the better the separation performance. From the contour plot, it can be seen that the pattern becomes progressively complicated as span angle increases. In the nominal center position, the region of good separation performance (the dark stripe) extends toward lower frequency limit (near 100 Hz) for the 120-deg span than the frequency limit (above 1 kHz) for the 10-deg span. On the other hand, the region of ringing frequencies (the white stripes for positive head displacements) occurs at lower frequency (600 Hz) for the 120-deg span versus 6 kHz for the 10-deg span. Thus, stereo dipole indeed has the advantage of having a much higher usable frequency limit before hitting the first ringing frequency which could lead to high gain inverse filters. However, it is argued by the authors that stereo dipole also suffers performance problems at low frequencies. These facts also suggest that large span arrangement should be used at low frequency, while small span arrangement should be used at high frequency, as suggested by many previous researchers.[11–16]

In order to explore further the effect of listening angle on the separation performance of CCS, an index, average channel separation, is defined as follows:

$$\frac{1}{M_2 - M_1 + 1} \sum_{k=M_1}^{M_2} 20 \times \log_{10}(|\mathrm{CHSP}_y(k)|) \text{ (dB)} \tag{15}$$

where M_1 and M_2 are the frequency indices of the lower and upper limits, and the subscript y denotes either L or R. In the simulation, the lower frequency limit was selected to be 100 Hz (M_1=60) below which the sound is known to be ineffective for localization. The average channel separation in relation to the listening angle and the lateral head displacement is shown with a contour plot in Fig. 6. Figures 6(a)–6(c) correspond to the average channel separations for three different frequency upper limits, 1 kHz (M_2=145), 6 kHz (M_2=211), and 20 kHz (M_2=255), re-

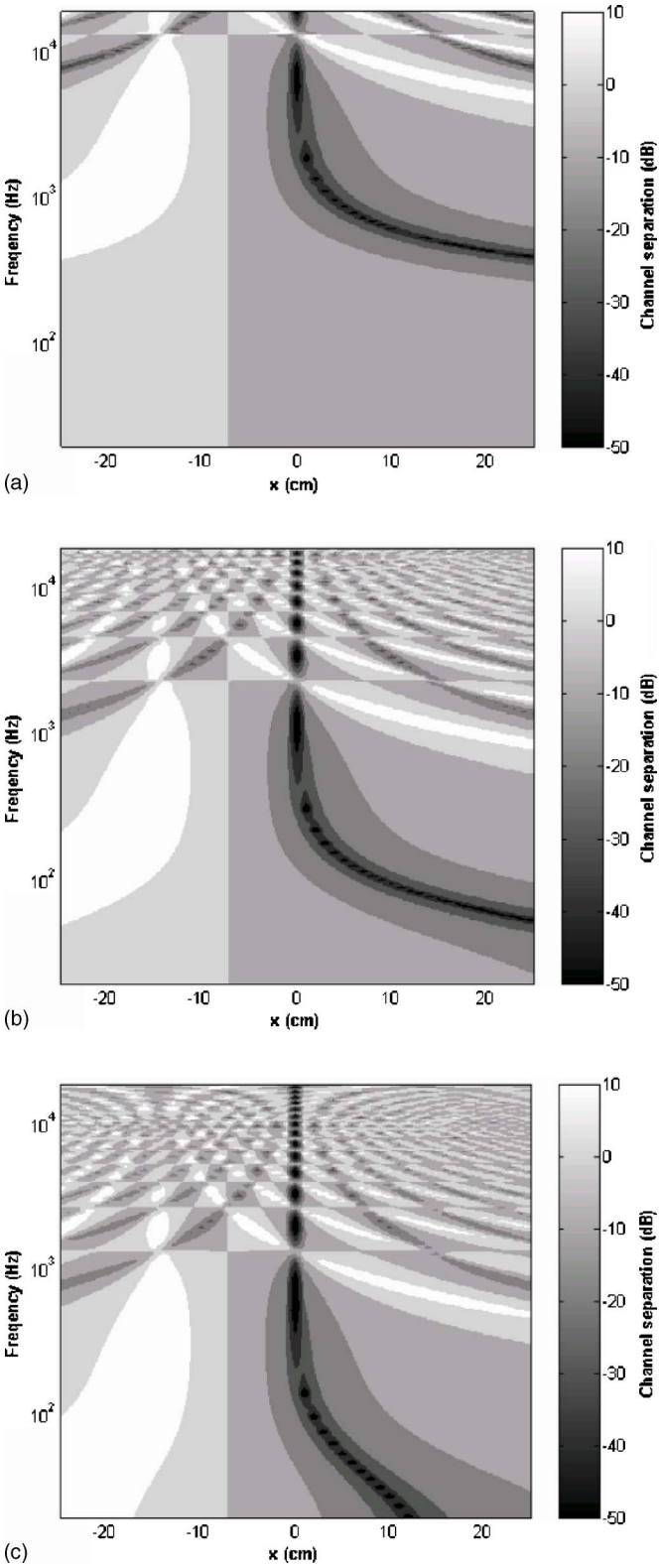

FIG. 5. The contour plots of channel separation at the right ear calculated using the point source model. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

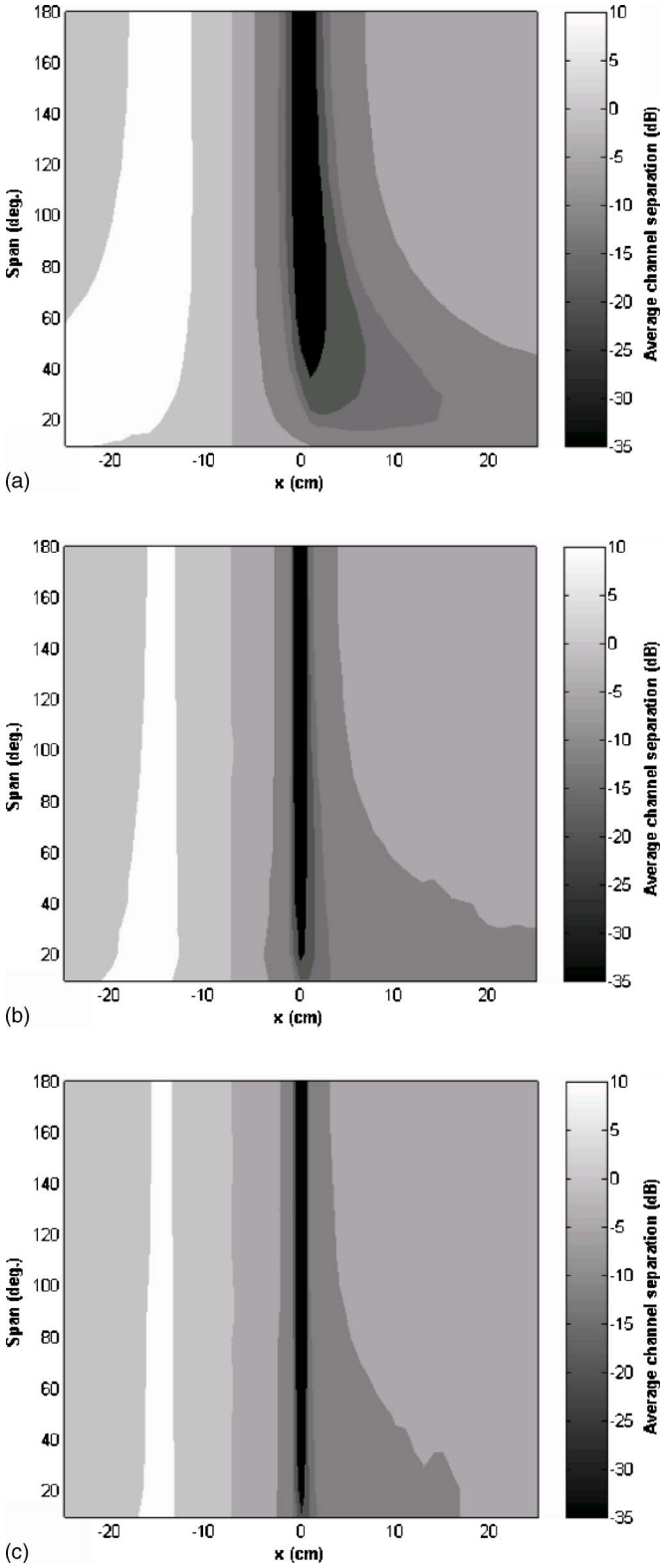

FIG. 6. The contour plots of average channel separation at the right ear calculated using the point source model. (a) Bandwidth to 1 kHz. (b) Bandwidth to 6 kHz. (c) Bandwidth to 20 kHz.

spectively. Using small span angle, a wider region of good separation performance (the second darkest stripe) can be attained at the expense of poor performance, especially for extremely small span. For example, Fig. 6(a) shows the 1-kHz-upper-limit average separation, where the lower tip of the second darkest region barely touches the 20-deg span.

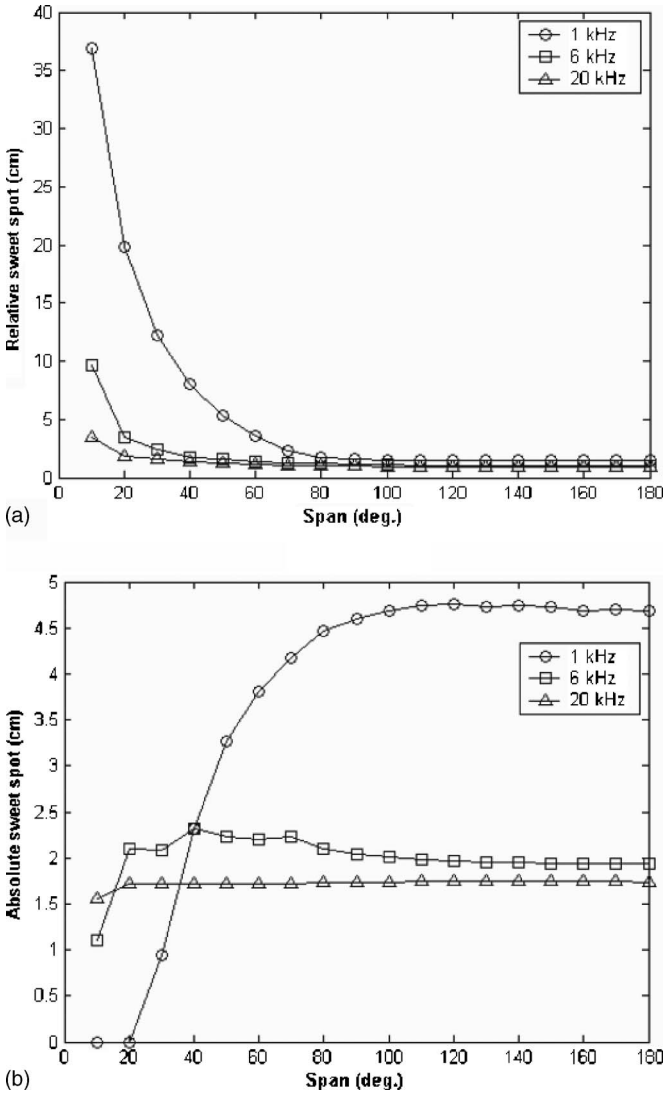

FIG. 7. Two sweet spot definitions calculated using the point source model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot.

The performance of CCS can also be characterized by sweet spot which refers to the region in which the CCS is effective. To be able to better assess the sweet spot quantitatively, two kinds of sweet spot are defined in the paper: the absolute sweet spot and the relative sweet spot. The size of absolute sweet spot is defined as two times the maximum leftward displacement that makes the average channel separation go below −12 dB. The size of relative sweet spot is defined with reference to Fig. 6 as two times the maximum leftward displacement for which the average channel separation is degraded by 12 dB as compared to that of the nominal center position ($x=0$). A value of −12 dB, or 25%, is an empirical value suggested by experience. For the absolute sweet spot, this value is the minimal requirement for CCS. For the relative sweet spot, this value corresponds to the point when the performance drops by 75% from the nominal position. The relative and absolute sweet spots calculated for the point source model are plotted versus span angle in Figs. 7(a) and 7(b), respectively. Three curves plotted in each figure correspond to three different bandwidths, 1, 6, and 20 kHz. As seen in the Fig. 7(a), the relative sweet spot is

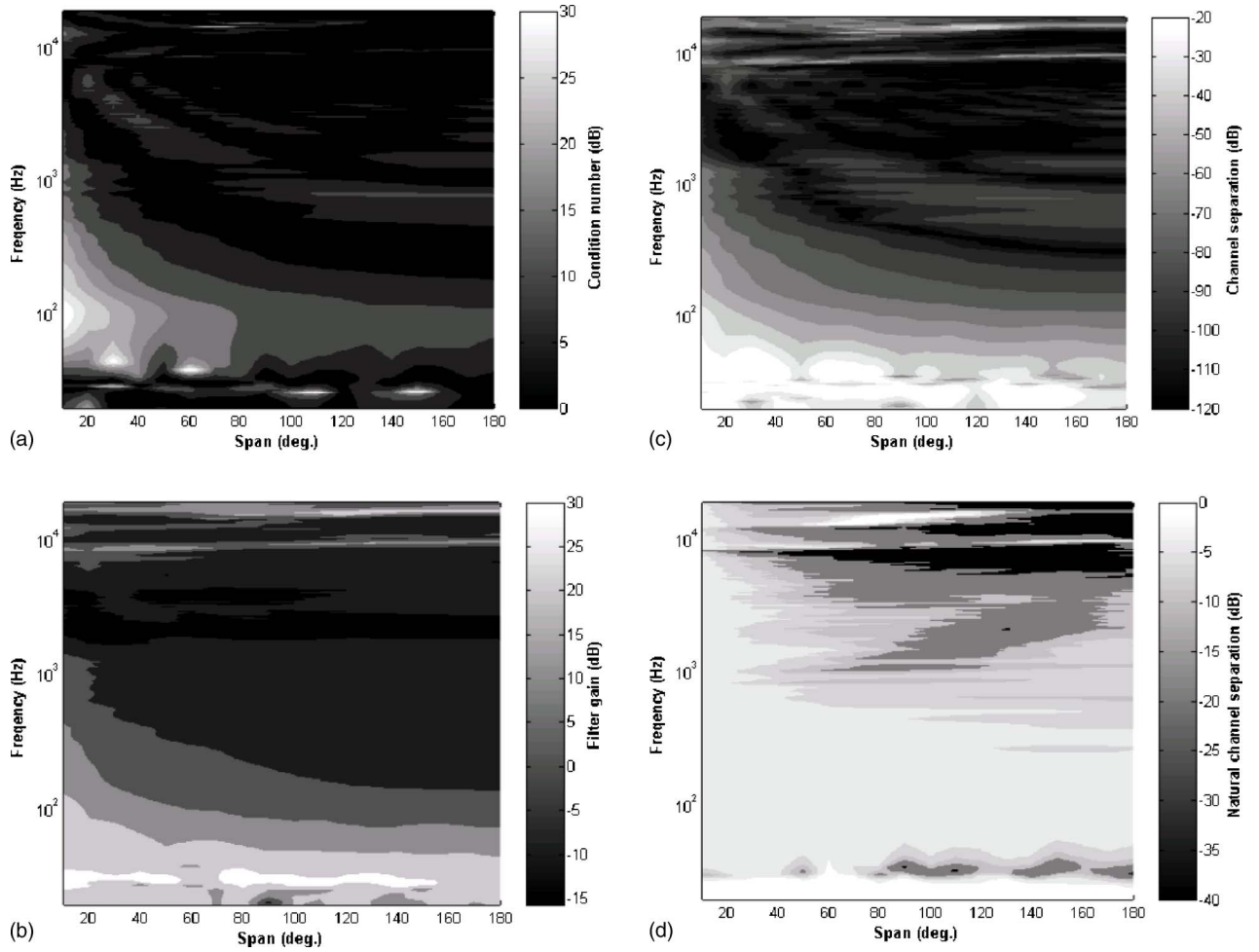

FIG. 8. The contour plots calculated using the HRTF model of (a) the condition number of acoustical plant matrix **H**, (b) the filter gain, (c) the channel separation, and (d) the uncompensated natural channel separation.

increased monotonically as the span is decreased, as predicted by previous researchers. This suggests that small span arrangement is more robust against head misalignment notwithstanding the poor separation performance at the nominal position. However, if the absolute sweet spot is taken as the robustness index, the conclusion is quite different. If this definition of sweet spot is used, the simulation result suggests that the optimal span angle ranges from 80 to 180 deg.

B. HRTF model

In addition to the point source model, a more sophisticated model based on HRTF is employed in the simulation to better account for the diffraction and shadowing effects due to the head, ears, and torso. The HRTF database measured by MIT Media Lab was employed. In the nominal position, the plant transfer function matrix is written as

$$\mathbf{H} = \begin{bmatrix} H_\theta^i & H_\theta^c \\ H_\theta^c & H_\theta^i \end{bmatrix}, \tag{16}$$

where θ is the span angle and the superscript i and c refer to ipsilateral and contralateral side, respectively. As the head moves to the right by x centimeters, the plant matrix is no longer symmetric and should be modified. The azimuth angle should be modified according to

$$\theta_{\mathrm{L}} = \tan^{-1}\frac{l\sin\theta_{\mathrm{L}_0} + x}{l\cos\theta_{\mathrm{L}_0}}, \tag{17a}$$

$$\theta_{\mathrm{R}} = \tan^{-1}\frac{l\sin\theta_{\mathrm{R}_0} - x}{l\cos\theta_{\mathrm{R}_0}}, \tag{17b}$$

where θ_{L_0} and θ_{R_0} are the angles in the nominal position, i.e., $x=0$. Linear interpolation is called for when the angle is not a multiple of a five-degree interval as the database was originally organized.[19] In addition to angles, the magnitudes and phases are also adjusted to account for attenuation and delay due to distance change. Thus,

$$\mathbf{H} = \begin{bmatrix} H_{\theta_{\mathrm{L}}}^i & H_{\theta_{\mathrm{R}}}^c \\ H_{\theta_{\mathrm{L}}}^c & H_{\theta_{\mathrm{R}}}^i \end{bmatrix} \times \begin{bmatrix} \frac{l_{\mathrm{LL}_0}}{l_{\mathrm{LL}}} e^{\frac{-j\omega(l_{\mathrm{LL}}-l_{\mathrm{LL}_0})}{c}} & \frac{l_{\mathrm{RL}_0}}{l_{\mathrm{RL}}} e^{\frac{-j\omega(l_{\mathrm{RL}}-l_{\mathrm{RL}_0})}{c}} \\ \frac{l_{\mathrm{LR}_0}}{l_{\mathrm{LR}}} e^{\frac{-j\omega(l_{\mathrm{LR}}-l_{\mathrm{LR}_0})}{c}} & \frac{l_{\mathrm{RR}_0}}{l_{\mathrm{RR}}} e^{\frac{-j\omega(l_{\mathrm{RR}}-l_{\mathrm{RR}_0})}{c}} \end{bmatrix}, \tag{18}$$

where the subscript "0" refers to the nominal position.

The contour plots of the condition number, filter gain, and channel separation are shown in Figs. 8(a)–8(c). The

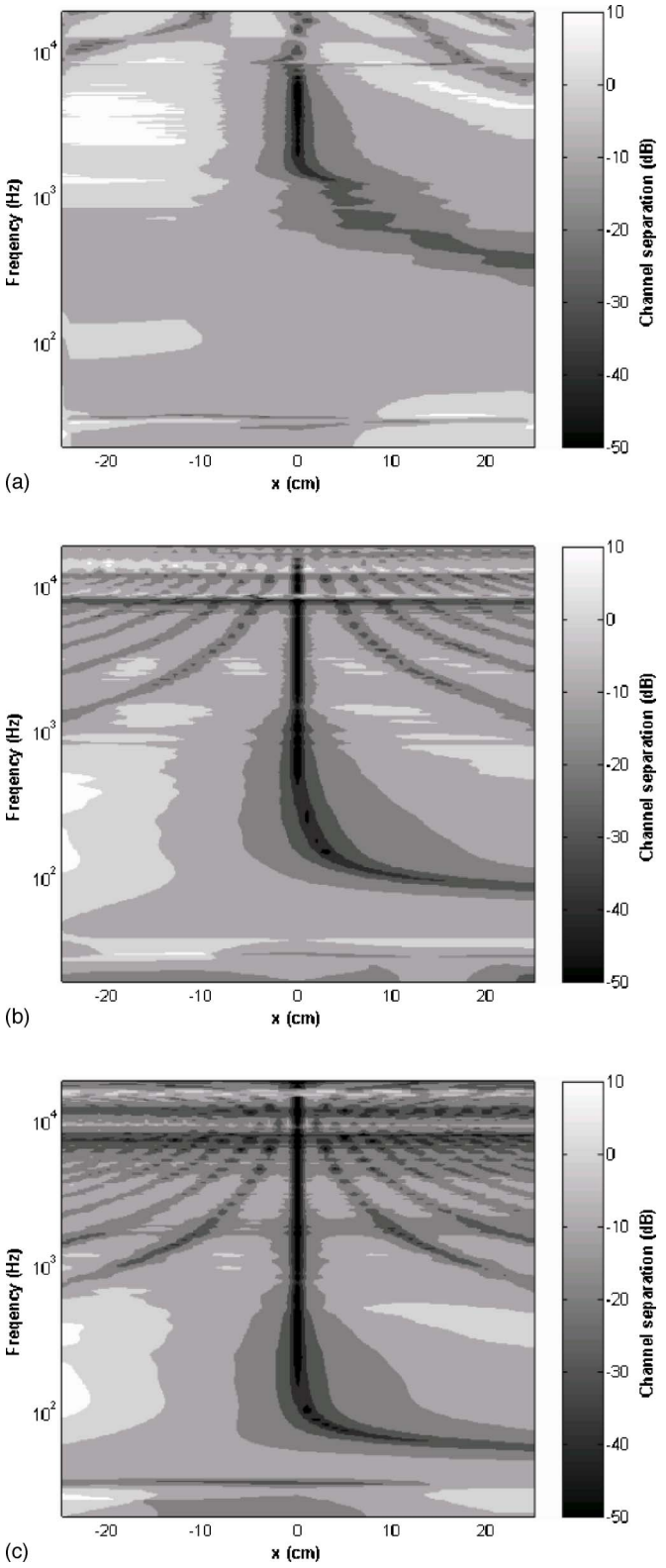

FIG. 9. The contour plots of channel separation measured at the right ear of the acoustic manikin. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

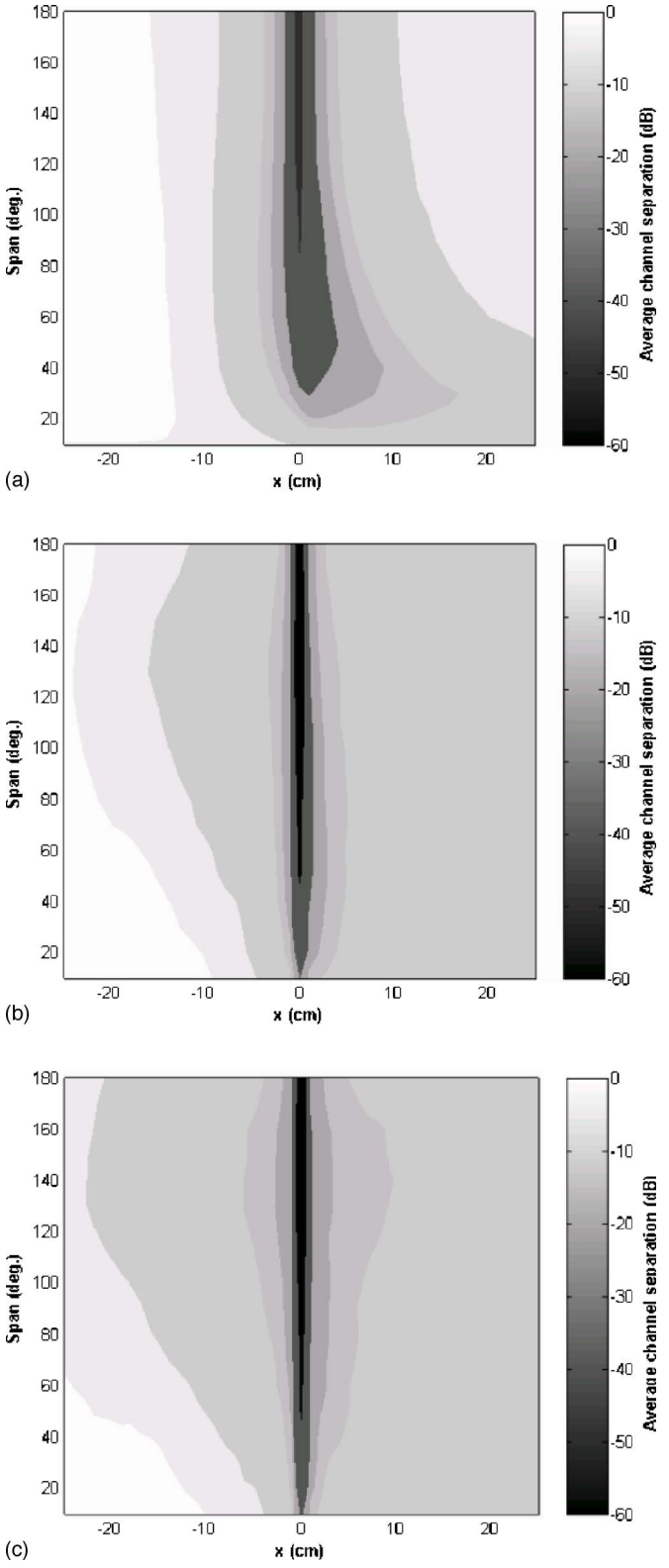

FIG. 10. The contour plots of band-average channel separation measured at the right ear of the acoustic manikin. (a) 1 kHz bandwidth. (b) 6 kHz bandwidth. (c) 20 kHz bandwidth.

uncompensated natural channel separation is also shown in Fig. 8(d) for reference, where the effect of head shadowing is clearly visible. By and large, the results of point source and HRTF follow a similar trend except one important distinction. Because of head shadowing at high frequencies, ringing does not show up in the HRTF results as pronouncedly as in the point source model except a constant ringing around 8–10 kHz due to the concha dip which is almost independent of span. The operation zone of HRTF is thus bounded from above by the concha dip, in contrast to the point source case that is bounded from above by the first ringing. This suggests that a large span arrangement seems to provide bet-

ter numerical stability with a larger useful frequency range than the small span arrangement. The separation performance at high frequencies for large spans is also better (reflected by more dark areas) than that of the small span owing to natural separation provided by head shadowing.

The contour plots of channel separation versus displacement and frequency are shown in Figs. 9(a)–9(c), corresponding to span angles 10, 60, and 120 deg, respectively. The trends of this result are largely the same as that of the point source model. The separation performance at low frequencies is still not good for the 10-deg span [Fig. 9(a)]. Figures 10(a)–10(c) show the contour plots of average channel separation versus displacement and span angle for frequency bandwidth, 1, 6, and 20 kHz, respectively. The trend of the HRTF result is similar to that of the point source result if only a narrow bandwidth, e.g., 1 kHz, is considered [Fig. 6(a) versus Fig. 10(a)]. However, if average separation performance is calculated for a larger bandwidth, e.g., 20 kHz, the results turn out to be quite different. The average performance is poor for extremely small spans. The region of good performance (the darkest strip) is mainly located around the median span area, say, from 100 to 160 deg. This difference of conclusion with the previous point source model is again due to the fact that the head shadowing effect will come into play at high frequencies.

The relative and absolute sweet spots, as defined previously in the point source simulation, are calculated for the HRTF model in three different bandwidths, 1, 6, and 20 kHz, as shown in Figs. 11(a) and 11(b). Similar to the point source results, the relative sweet spot is increased monotonically as the span is decreased, which suggests that small span arrangement is relatively robust against head misalignment notwithstanding the poor separation performance at the nominal position. On the other hand, the results of the absolute sweet spot suggest that arrangements with listening angles ranging from 120 to 150 deg [the intersection of bandwidth of 6 and 20 kHz in Fig 11(b)] seem to be good choices.

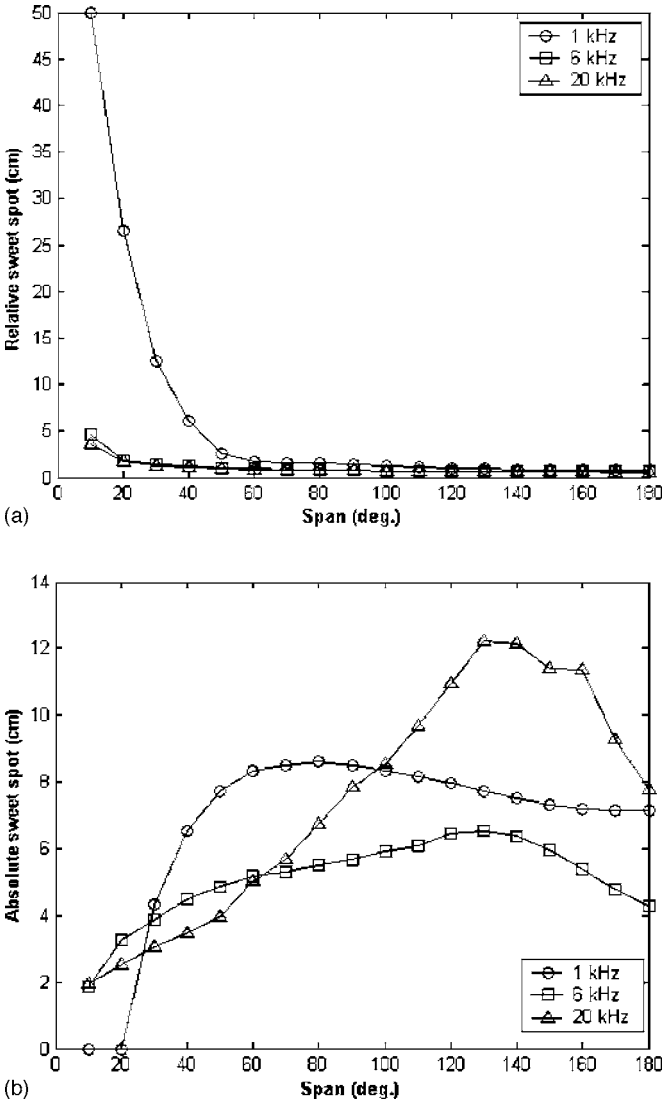

FIG. 11. Two sweet spot definitions calculated using the HRTF model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot.

IV. OBJECTIVE AND SUBJECTIVE EXPERIMENTS

The forgoing simulation results suggest that the optimal listening angle ranges from 120 to 150 deg. This observation is further examined in a series of objective and subjective experiments. Three loudspeaker arrangements with 10-, 60-, and 120-deg spans were compared in the experiments. The 10-deg span represents stereo dipole. The 60-deg span is suggested in the ITU standard of a multichannel stereophonic system.[27] The 120-deg span represents the optimal span previously found in the simulation. All experiments were carried out in an anechoic room, as shown in Fig. 12.

A. Objective experiment

This experiment employed a 5.1-channel loudspeaker system, Inspire 5.1 5300 of Creative, and a digital signal processor (DSP), Blackfin-533, of Analog Device. The microphones and the preamplifier used are GRAS 40AC and GRAS 26AM. The plant transfer function matrixes were measured on an acoustical manikin, KEMAR (Knowles Electronics Manikin for Acoustic Research) along with the ear model, DB-065.

The designed CCS filters were implemented on the DSP using 512-tapped Finite Impulse Response (FlR) filters. The performance of CCS was evaluated in terms of channel sepa-

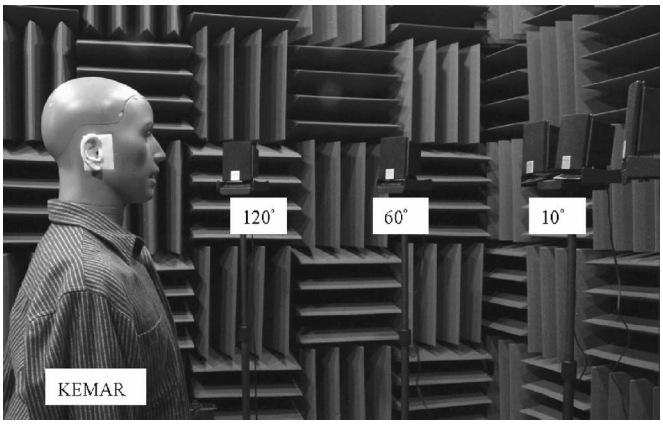

FIG. 12. Photo of the experimental arrangement.

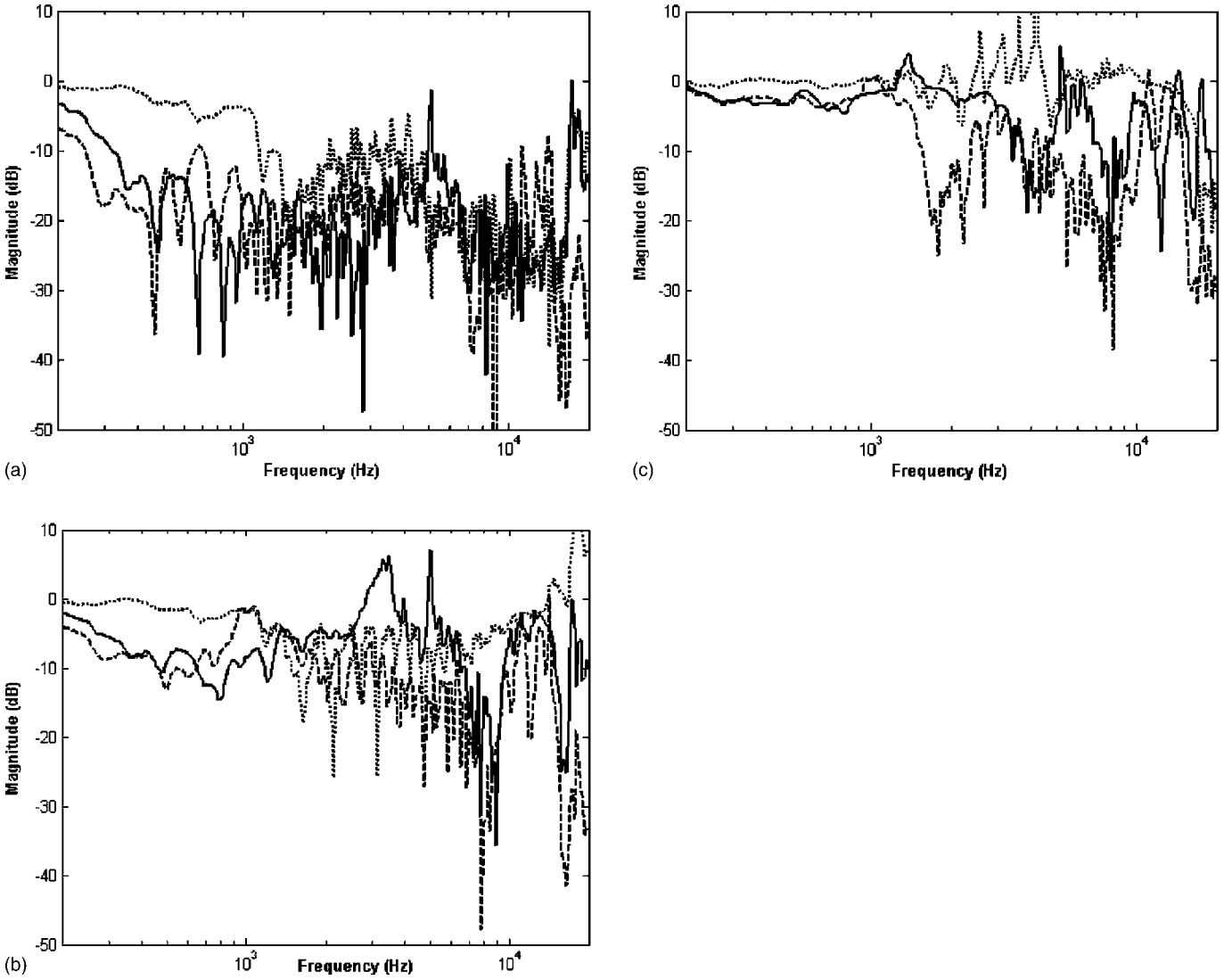

FIG. 13. Channel separations measured at the right ear of the acoustic manikin. The dotted lines, solid lines, and dashed lines represent 10-, 60-, and 120-deg spans, respectively. (a) In the nominal position (x=0 cm). (b) Rightward 5 cm displacement. (c) Rightward 10 cm displacement.

ration. Figure 13(a) shows the right-ear channel separation at the nominal position with three span angles. The x axis and y axis represent frequency in hertz and magnitude in decibels, respectively. The dotted line, the solid line, and the dashed line signify 10°, 60°, and 120° span angles, respectively. The results of Figs. 13(b) and 13(c) were obtained for the cases when the manikin was moved to the right by 5 and 10 cm. Notable of these results is that the 10-deg span performed badly at the frequencies below 1 kHz. The separation performance significantly degraded by as much as 15 dB as the head moved to the right by 5 cm irrespective of which span was used. As the head was displaced by 10 cm, CCS failed almost completely, except at high frequencies, when the large 120-deg span arrangement still maintained natural separation because of head shadowing.

B. Subjective experiment

For the purpose of comparing the CCS with different span angles, a subjective listening experiment of source localization was undertaken in the anechoic room. Eleven subjects participated in the test. The listeners were instructed to sit at three positions: the nominal position, 5-cm displacement to the right, and 10-cm displacement to the right. In order to ensure that each listener sat at the same designated position, the test subjects were asked to rest their chins on a steel frame. The height of the listener's ear was 120 cm which is the same height as the loudspeaker. A pink noise was used as the test stimulus whose bandwidth ranges from 20 Hz to 20 kHz and the reproduction level was 95 dB. Each stimulus was played five times in 25-ms duration with 50-ms silent interval. Virtual sound images at 12 prespecified directions on the horizontal plane with increment 30° azimuth are rendered by using HRTFs. Listeners were well trained by playing the stimuli of all angles prior to the test. The listeners were asked to report the perceived direction of source in the range (−180,180] with a 30-deg interval. Experiments were divided into two groups: 10 deg versus 120 deg and 60 deg versus 120 deg. The experiments were blind tests in that stimuli were played randomly without informing the subjects the source direction. One session of test lasts 15–20 min.

The results of the localization test are shown in terms of target angles versus judged angles in Figs. 14–16, corre-

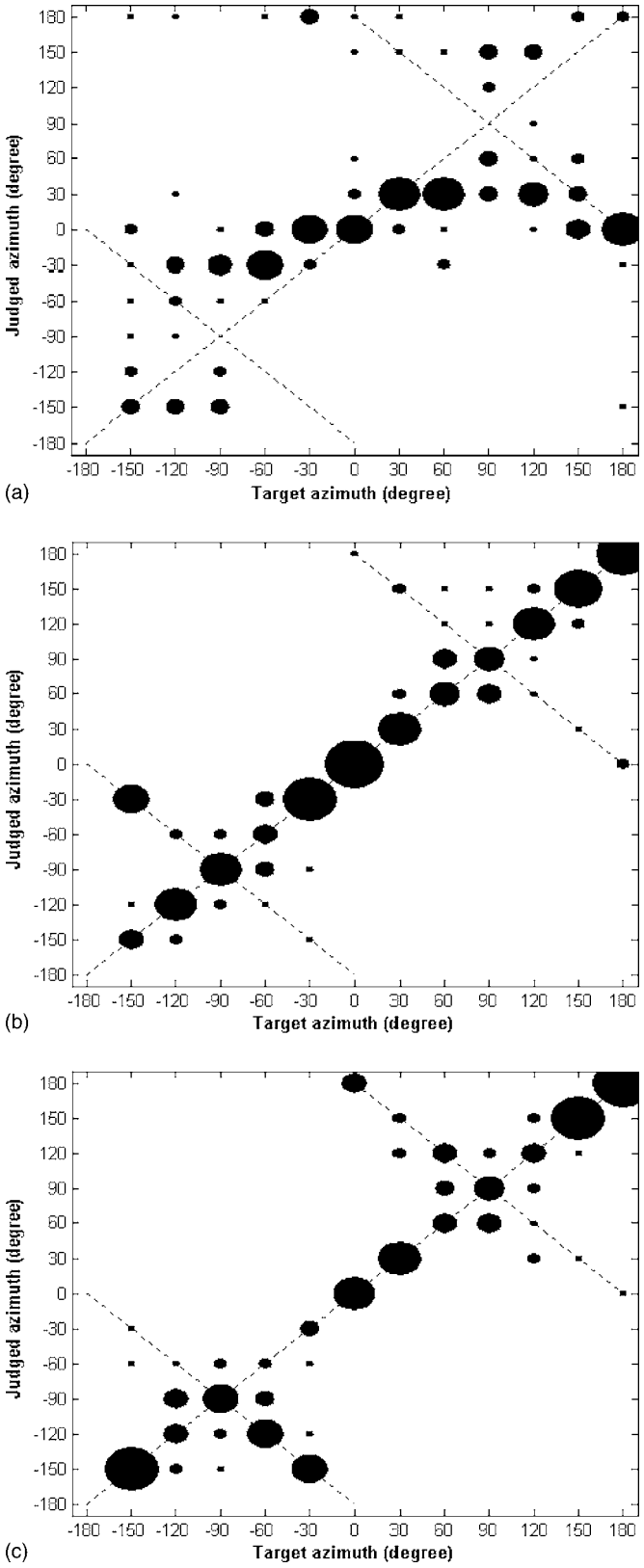

FIG. 14. Results of the subjective localization test of azimuth angles with no head displacement. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

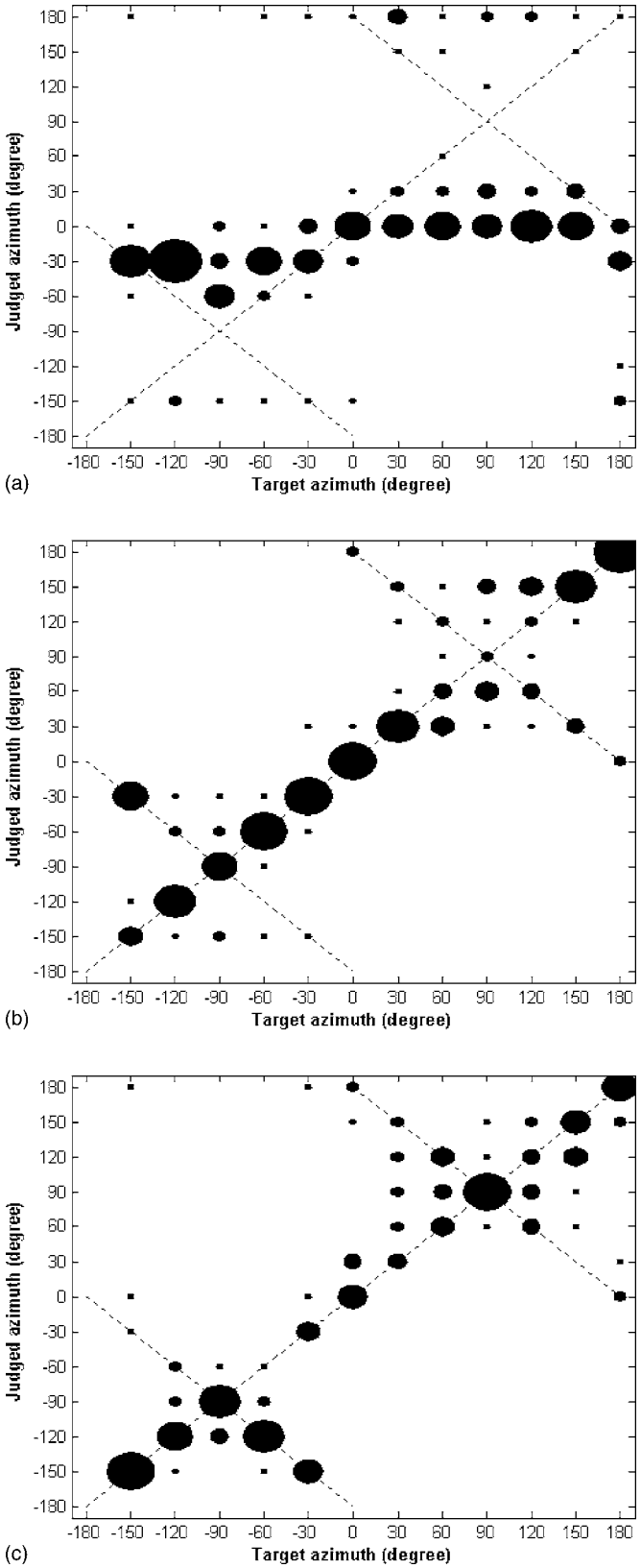

FIG. 15. Results of the subjective localization test of azimuth angles with 5-cm head displacement to the right. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

sponding to the cases of nominal position, 5-cm displacement to the right, and 10-cm displacement to the right. In each figure, subplot (a) to (c) refer to the 10°, 60°, and 120° spans, respectively. The size of each circle is proportional to the number of the listeners who localized the same perceived angle. The 45-deg line indicates the perfect localization. It is observed from the results that the subjects tend to localize the sources within ±30 deg about the center line using the 10-deg span arrangement, especially when there is head dis-

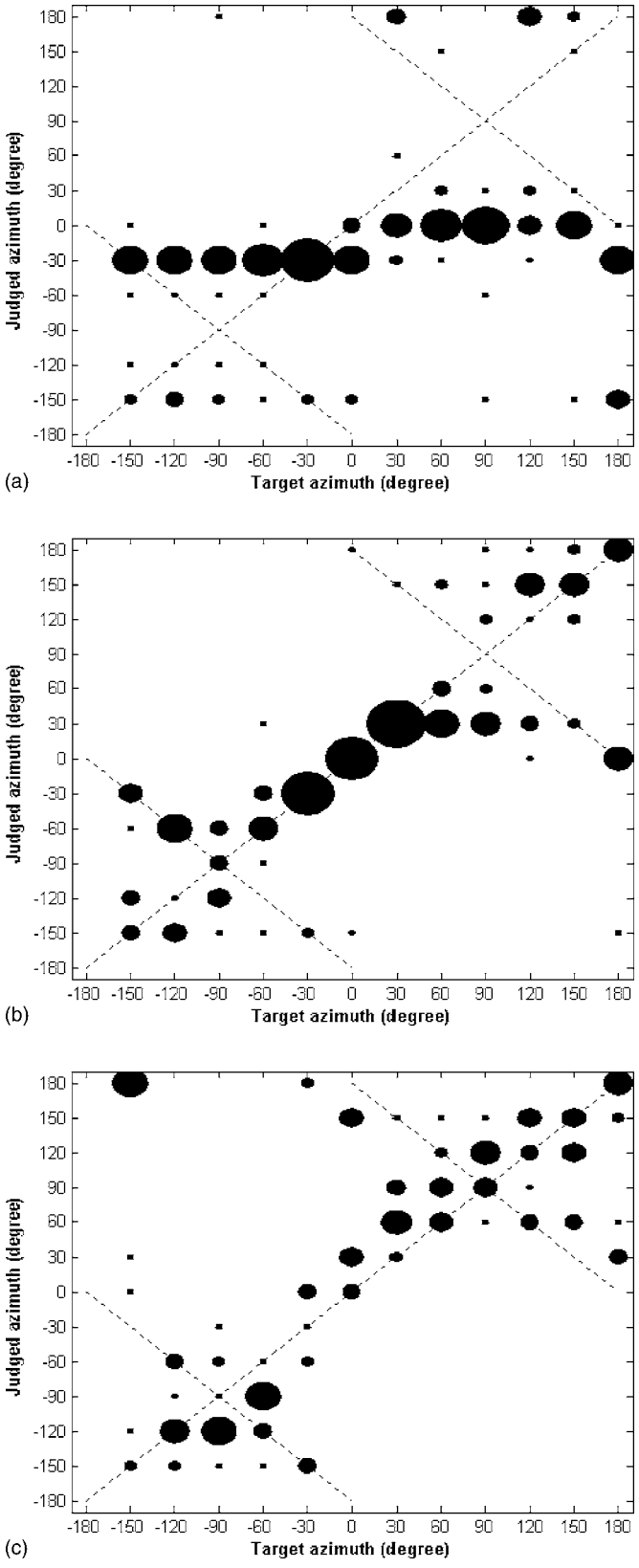

FIG. 16. Results of the subjective localization test of azimuth angles with 10-cm head displacement to the right. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

placement. On the other hand, the 60-deg span and the 120-deg span were found to be effective in localizing good frontal images and rear images albeit some front-back reversals. Localization error increases with head displacement irrespective of which span arrangement was used. The 10-deg span seemed to have difficulty localizing sources outside the subtending angle because the separation performance in low frequencies is too poor in small span arrangement to maintain proper spatial cues such as interaural time difference (ITD) which works only under 1 kHz. In contrast, the arrangement with large span appears to be more robust than the small span because head shadowing and panning effect help to provide localization effect to certain degree even if CCS breaks down.

TABLE I. The description of five levels of grade for the subjective localization test.

Grade	Description
5	The judged angle is the same as the target angle
4	30° difference between the judged angle and the target angle
3	Front-back reversal of the judged angle identical to the target angle
2	30° difference between front-back reversal of the judged angle and the target angle
1	Otherwise

To justify the finding, a Friedman test on the subjective localization results in relation to span was conducted. These results were preprocessed into five levels of grade, as described in Table I. The results of the Friedman test are summarized in Table II for the first and second groups. Figure 17(a) shows the medians (small square), quartiles (box), and ranges (whiskers) of the 10-deg span and the 120-deg span. Friedman test output of the first group in Table II reveals that the span effect is statistically significant ($p<0.001$). This indicted that the 120-deg span outperformed the 10-deg span. Figure 17(b) shows the medians, quartiles, and ranges of the 60-deg span and the 120-deg span. The Friedman test of the second group in Table II reveals that the difference of performance of two listening angles is found statistically insignificant ($p<0.3458$). This does not seem to agree with the prediction of the previous simulation that the 120-deg span should perform slightly better than the 60-deg span. It is suspected that the enormous span of 120-deg arrangement is actually quite detrimental to localizing sources at the center position, especially when CCS beaks down. Experience shows overly large angle arrangements seem to have difficulties in positioning images at the center region. In fact, some of the test subjects reported that it sounded like there was an opening of sound field in the front. This offsets somewhat the expected performance gain of CCS using large span arrangement.

TABLE II. The Friedman test result of the subjective experiments.

	First group (10 vs 120)	Second group (60 vs 120)
Chi-Squares ($N=396$, $df=1$)	47.4568	0.8889
Significant p value	<0.001	<0.3458

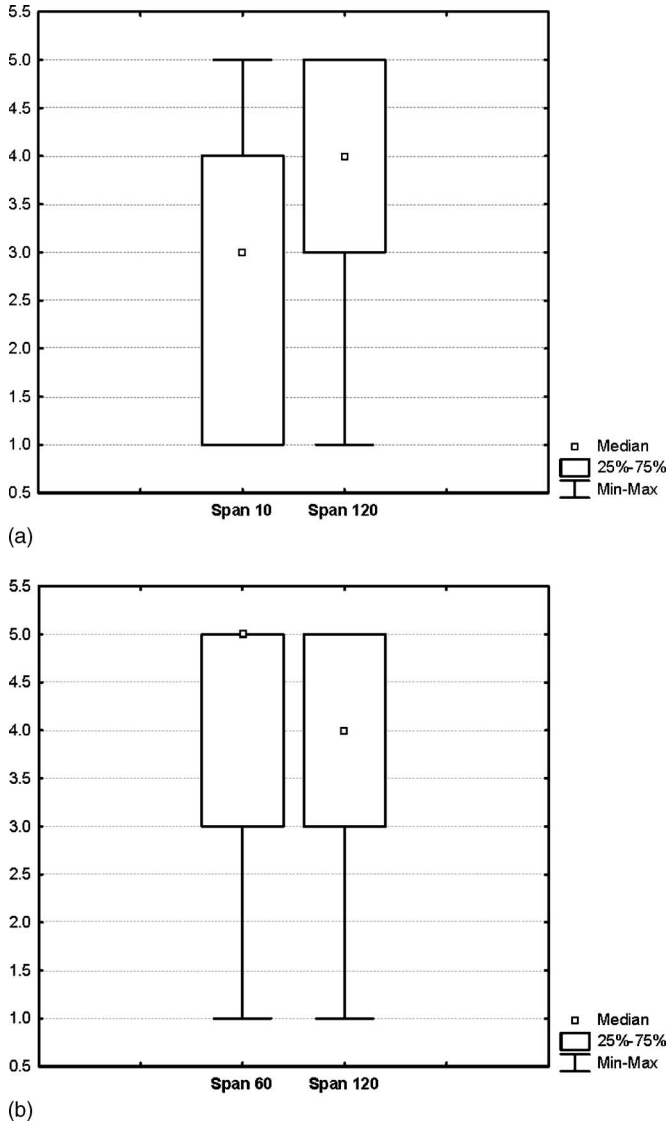

FIG. 17. The box and whisker plots, (a) 10-deg arrangement vs 120-deg arrangement. (b) 60-deg arrangement vs 120-deg arrangement.

V. CONCLUSIONS

A comprehensive study has been conducted to explore the effects of listening angle on crosstalk cancellation in spatial sound reproduction using two-channel stereo systems. The intention is to establish a sustainable configuration of CCS that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice. Similar to the previous research which focuses mainly on numerical stability, the present work arrives at the conclusion that inversion of ill-conditioned systems results in high gain filters, loss of dynamic range, and hence separation performance. Regularization is required to compromise between numerical stability and separation performance. However, findings different from the previous study had also been reached because this work employed a comprehensive approach. First, it is found from the HRTF results that the problem of high frequency ringing is not as critical as in the point source model owing to head shadowing. In addition, poor conditioning, high gain, and low performance problems at low frequencies may arise for extremely small span arrangements, whereas there is broader useful frequency range with performance and numerical stability if wide span arrangement can be used. The effects of listening angle were also examined in the context of the sweet spot. Two kinds of sweet spot definitions are employed in the simulation. The relative sweet spot suggests that robustness is excellent with the use of small span arrangement notwithstanding the poor performance in the nominal position, which is in agreement with the previous research. However, it is not very useful in practical application if the average channel separation in the sweet spot is very poor even though it is relatively robust. Therefore, in addition to the conventional relative definition, we suggest another definition, the *absolute* sweet spot, to make the evaluation more complete. In an absolute sweet spot, the performance is guaranteed in complement to the relative robustness, which is desirable in practical use of the CCS. The results of absolute sweet spot reveal that arrangements with a listening angle ranging from 120 to 150 deg are optimal choices.

To justify the conjectures above, objective and subjective experiments were undertaken in an anechoic room for three loudspeaker arrangements, including the stereo dipole (10 deg), standard span (60 deg), and proposed span (120 deg). The results postprocessed by the Friedman test indicate that the 120-deg configuration performs comparably well as the standard 60-deg configuration, but is better than the 10-deg configuration. Small span arrangement produces a large relative sweet spot because head displacement would cause minimal change of time-of-arrival differences between two loudspeakers using closely spaced loudspeakers. This configuration is well suited to applications that must be spatially compact, e.g., mobile phones and other portable devices. Nevertheless, the benefit of small span arrangement comes at the price of poor conditioning, high gain, and limited performance problems at low frequencies. Apart from this, due to the lack of natural high frequency separation provided by head shadowing, the small span arrangement is not able to position "out-of-range" source when CCS breaks down at high frequencies, where the phantom source is incorrectly panned within a narrow span. The arrangement with large span appears to be more effective than the small span because head shadowing and panning effect help to provide a localization effect to a certain degree even if CCS breaks down. While it may seem from this report that large-span configuration is predominantly favored, problems inherent to large span prevent the span to grow indefinitely, e.g., sound image stability will become an issue for wide apart loudspeakers. A practical recommendation is perhaps the conventional 60-deg configuration which is a reasonable compromise between the two extremes (10 and 120 deg) to achieve both robustness and performance. It was also found that the 120-deg arrangement did not perform as well as the 60-deg arrangement in positioning frontal images. If an additional center loudspeaker is available, the 3/0 format with 120-deg span would be an ideal choice.

ACKNOWLEDGMENTS

The work was supported by the National Science Council in Taiwan, Republic of China, under the Project No. NSC94-2212-E009-019.

[1]J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).
[2]D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (AP Professional, Cambridge, MA, 1994).
[3]R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," IEEE Int. Convention Record **7**, 150–155 (1963).
[4]P. Damaske and V. Mellert, "A procedure for generating directionally accurate sound images in the upper-half space using two loudspeakers," Acoustics **22**, 154–162 (1969).
[5]D. H. Cooper, "Calculator program for head-related transfer functions," J. Audio Eng. Soc. **30**, 34–38 (1982).
[6]W. G. Gardner, "Transaural 3D audio," MIT Media Laboratory Tech. Report 342 (1995).
[7]D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," J. Audio Eng. Soc. **37**, 3–19 (1989).
[8]J. L. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," J. Audio Eng. Soc. **44**, 683–705 (1996).
[9]C. Kyriakakis, T. Holman, J. S. Lim, H. Homg, and H. Neven, "Signal processing, acoustics, and psychoacoustics for high-quality desktop audio," J. Visual Commun. Image Represent **9**, 51–61 (1997).
[10]C. Kyriakakis, "Fundamental and technological limitations of immersive audio systems," IEEE Signal Process. Mag. **86**, 941–951 (1998).
[11]D. B. Ward and G. W. Elko, "Optimal loudspeaker spacing for robust crosstalk cancellation," Proc. ICASSP 98 (IEEE, Seattle, WA, 1998), pp. 3541–3544.
[12]D. B. Ward and G. W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," IEEE Signal Process. Lett. **6**(5), 106–108 (1999).
[13]D. B. Ward, "Joint squares optimization for robust acoustic crosstalk cancellation," IEEE Trans. Speech Audio Process. **8**(2), 211–215 (2000).
[14]O. Kirkeby, P. A. Nelson, and H. Hamada, "The "stereo dipole" a virtual source imaging system using two closely spaced loudspeakers," J. Audio Eng. Soc. **46**, 387–395 (1998).
[15]T. Takeuchi and P. A. Nelson, "Robustness to head misalignment of virtual sound imaging systems," J. Audio Eng. Soc. **109**, 958–971 (2001).
[16]T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," J. Acoust. Soc. Am. **112**, 2786–2797 (2002).
[17]P. A. Nelson and J. F. W. Rose, "Errors in two-point sound reproduction," J. Acoust. Soc. Am. **118**(1), 193–204, 2005.
[18]M. R. Bai, C. W. Tung, and C. C. Lee, "Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm," J. Acoust. Soc. Am. **117**, 2802–2813 (2005).
[19]W. G. Gardner, *3-D Audio Using Loudspeakers* (Kluwer Academic, Dordrecht, 1998).
[20]W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908 (1995).
[21]W. G. Gardner and K. D. Martin, *KEMAR HRTF Measurements* (MIT's Media Lab, http://sound.media.mit.edu/KEMAR.html, 1994).
[22]B. Noble, *Applied Linear Algebra* (Prentice-Hall, Englewoods, NJ, 1988).
[23]O. Kirkeby, P. A. Nelson, and H. Hamada, "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**, 189–194 (1998).
[24]A. Schuhmacher and J. Hald, "Sound source reconstruction using inverse boundary element calculations," J. Acoust. Soc. Am. **113**, 114–127 (2003).
[25]M. R. Bai and C. C. Lee, "Development and implementation of cross-talk cancellation system in spatial audio reproduction based on the subband filtering," J. Sound Vib. **290**, 1269–1289 (2006).
[26]V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Platz, New York, 2001).
[27]ITU-R Rec. BS.775–1, *Multi-Channel Stereonhonic Sound System With or Without Accompanying Picture* (International Telecommunications Union, Geneva, Switzerland, 1992–1994).