Channel separation of crosstalk cancellation systems with mismatched and misaligned sound sources

Xiaojun Qiu[a)]
Key Laboratory of Modern Acoustics, Institute of Acoustics, Nanjing University, Jiangsu 210093, China

Bruno Masiero and Michael Vorländer
Institute of Technical Acoustics, RWTH Aachen University, D-52056 Aachen, Germany

(Received 7 October 2008; revised 25 July 2009; accepted 28 July 2009)

Loudspeakers in virtual sound imaging systems are usually modeled as omnidirectional monopole sources. These models are, however, only an approximation for the low frequency range. This paper presents an analytical model of crosstalk cancellation systems in a free field which takes into account the scattering and spatial characteristics of the sound sources. Based on the proposed model, the effects caused by the spatial characteristics of the sound source and its misalignments on the performance of the crosstalk cancellation system are studied numerically. It is found that although the factors such as the directivity of the sound sources and the distance between the sound sources and receiver affect the performance of the system to a certain extent, the channel separation of the crosstalk cancellation system, however, is most sensitive to the misalignment of the subtended angle of the sound sources. Therefore, if highly accurate binaural cues are required in practical applications, the type and characteristics of the playback sound sources, their locations, and orientations all should be considered carefully.
 [DOI: 10.1121/1.3206660]

PACS number(s): 43.38.Md, 43.38.Vk, 43.66.Pn [AJZ] Pages: 1796–1806

I. INTRODUCTION

During the past few decades, considerable attention has been paid to the area of sound field reproduction, raising many issues in acoustics and signal processing. Several approaches can be used to reproduce a required sound with loudspeakers, from which the most popular are the binaural techniques, ambisonics, and wave field synthesis.[1,2] If the sound is required to be reproduced at only two points in space (the ears of a listener), two loudspeakers can be used to generate a virtual sound image by pre-filtering the sound signal with a pair of causal linear filters.[2–6] If the sound arrival directions at the center point of a volume are of most interest, an ambisonics system with a small number of loudspeakers can be used, which is based on spherical harmonic decomposition of the acoustic field centered at the listener position.[7] For even larger listening area, the wave field synthesis concept based on the Kirchhoff–Helmholtz integral equation and the high order ambisonics technique can be used.[7–9] This paper concentrates on the particular binaural virtual sound imaging system that uses two loudspeakers together with the crosstalk cancellation (CTC) technique.

The head related transfer functions (HRTFs) describe the filtering of the sound from a specific location by the diffraction and reflection properties of the head, pinna, and torso of an individual before it reaches the individual's eardrums.[1–4] A synthesized binaural signal can be created by convolving a sound with the appropriate HRTFs for generating spatial audio effects by using headphones or loudspeakers. The difficulty with the loudspeaker binaural virtual sound imaging system is the crosstalk in the contralateral paths from the loudspeakers to the listener's ears, which destroys the three-dimensional (3D) cues of the binaural signal, and the CTC technique is the most commonly used way to solve the problem.

CTC systems (the systems that use CTC technique) have been studied for nearly 50 years, and an in-depth review of relevant literature can be found in recent book and papers.[1–4] It has been observed that a nearly perfect CTC (greater than 40 dB up to 5000 Hz) can be obtained for a matched CTC system where the playback HRTFs are the same as the setup HRTFs used to design the CTC filters. However, for a mismatched system, the channel separation drops significantly. It has been concluded by Akeroyd *et al.*[3] that although test trials with CTC systems are very impressive and provide compelling subjective test results on the angle perception, only a matched system can yield sufficiently accurate binaural cues required for spatial hearing experiments.

There are many factors that make the playback HRTFs mismatch the setup ones. For instance, the difference between HRTFs of different individuals whose pinnae, heads, and torsos differ in size and shape, the misalignment of the listener's head position and orientation, the misalignment of position and orientation of playback sound sources, the variation in the spatial and temporal characteristics of the playback sound sources, and the variation in the playback acoustics environment caused by wall reflections. The loss in performance caused by the difference between the setup HRTFs and the playback HRTFs of different individuals has been studied by Akeroyd *et al.*[3] and it turned out that the average channel separation can decrease by up to 20 dB in

[a)]Author to whom correspondence should be addressed. Electronic mail: xjqiu@nju.edu.cn

some cases. The study of the robustness to head misalignment by Takeushi *et al.*[5] shows that a system with loudspeakers that are positioned closer to each other is more robust to the misalignment of the listener's head. Rose *et al.*[6] further analyzed this fact in terms of the off-axis asymmetric listener locations and found that the width of tolerable lateral head translations for asymmetric locations is comparable to that for the symmetric case. A recent study carried out by Bai *et al.*[4] shows that, despite the fact that using a small source span angle of 10° results in a large relative sweet spot, using a larger source span of 60° or even 120° is more desirable in practical applications. This can produce a larger absolute sweet spot where the channel separation performance is guaranteed.

In addition to these robustness studies which focused mainly on the listener, there are also several studies on the influence of playback environments. Ward[10] developed a closed-form expression to predict the performance of a CTC system in a reverberant environment. Subjective tests carried out by Lentz show that adding reflecting walls to the listening environment decreases the performance of the CTC system significantly.[11] In some practical applications, such as spatial hearing experiments, where highly accurate binaural cues are required, it becomes evident that the scattering and reflections from the loudspeaker cabinets have to be taken into account.[3] However, little is known about effects of mismatched playback sound sources on the CTC system. In the design of the CTC filters, the setup HRTFs used are usually measured in prior with an ideal small sound source which radiates equally in all direction. But the sound sources used in playback are usually not the same ones, and their time and spatial characteristics might be different. These mismatches might deteriorate the performance of a CTC system.

In sometime applications, for example, in a cave automatic virtual environment (CAVE), the directivity of the sound sources needs to be considered. A CAVE is an immersive virtual reality environment where video projectors are directed to three, four, five, or six of the walls of a room-sized cube.[11] In a CAVE, a user can walk around to have a proper view of 3D objects from different directions by wearing special glasses, so a dynamic CTC system has to be used to provide a valid CTC filter set for each position the user might be to generate proper virtual sound images. The term "dynamic" here means updating the CTC filter dynamically (adaptively) with the listener position change detected by a head tracker. Under this situation, the relative angle between the loudspeakers' orientation and the listener is not fixed and it might not be easy to align the orientation of the playback sound sources with the moving listener adequately in practice, so the directivity of the sound sources might need to be taken into account to achieve the required performance.[11]

All these facts motivate the current research to investigate the significance of the effects caused by the spatial characteristics and misalignment of playback sound sources. Loudspeakers used in CTC systems are usually considered as monopoles in existing models, but this model is only an approximation in the low frequency range. This paper develops first of all an analytical model of CTC systems in a free field, which takes into account the scattering and spatial characteristics of the sound sources. Then, based on the proposed model, the effects of the spatial characteristics and misalignments of sound sources on the performance of CTC systems are studied.

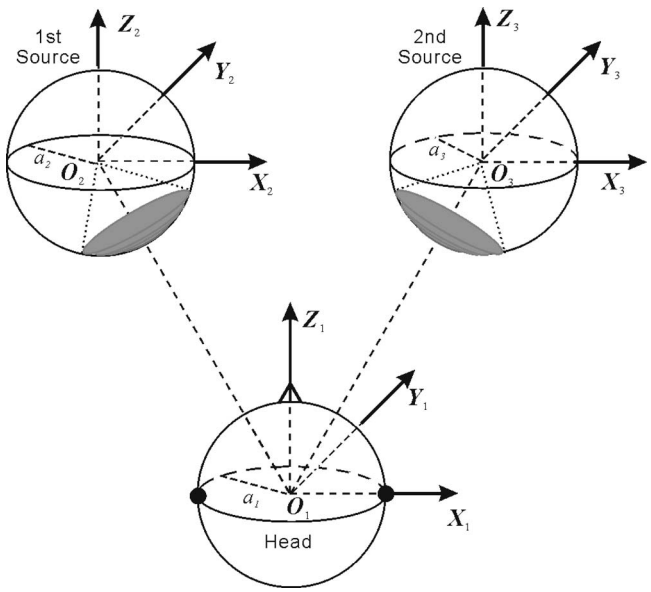

FIG. 1. The free field geometrical arrangements of the CTC systems with two vibrating cap sound sources on two rigid spheres and two receiving points on a third rigid sphere.

II. ANALYTICAL MODEL

A CTC system is modeled with three rigid spheres with a radius of a_i $(i=1,2,3)$, as shown in Fig. 1. Each sphere has its own spherical coordinate system referred to as O_i $(i=1,2,3)$. The coordinate systems O_2 and O_3 are obtained by a translational movement of the original spherical center from $\mathbf{r}_{o1}=(0,0,0)$ to $\mathbf{r}_{o2}=(d_{o2},\theta_{o2},\phi_{o2})$ or $\mathbf{r}_{o3}=(d_{o3},\theta_{o3},\phi_{o3})$ in the coordinate system O_1. The same receiving point is called $\mathbf{r}_{1r}=(r_{1r},\theta_{1r},\phi_{1r})$ in the coordinate system O_1 or $\mathbf{r}_{2r}=(r_{2r},\theta_{2r},\phi_{2r})$ in the coordinate system O_2 or $\mathbf{r}_{3r}=(r_{3r},\theta_{3r},\phi_{3r})$ in the coordinate system O_3.

Sphere 1 is used to model the head of a listener and Spheres 2 and 3 are used to approximate the loudspeaker cabinets. The model for the sound source used here consists of a rigid sphere with a radially vibrating cap (gray part shown in Fig. 1) defined by a certain solid angle. It is based on the assumption that the diffraction effects caused by a sphere and a cube with similar dimensions are similar if the dimensions of the diffracting bodies are considerably smaller than the acoustic wavelength.[12] Despite the fact that the error caused by this idealization increases in higher frequency, it can be used to show the general effects of the loudspeaker's directional proprieties and diffractions on the performance of the CTC systems and provides some insights into the basic mechanisms.

The simplest Z axis symmetrical case is considered first to calculate the spherical harmonic coefficients of a cap source in its own coordinate system. The velocity distribution of a cap source located at the North Pole on the surface of such a sphere is assumed as

$$U_o(\theta,\phi)=\begin{cases}u_0, & 0\le\theta\le\theta_s, 0\le\phi<2\pi\\ 0, & \theta_s<\theta\le\pi, 0\le\phi<2\pi,\end{cases} \tag{1}$$

where θ_s is half of the cap source span angle, and the time-dependent factor $e^{j\omega t}$ is suppressed throughout the analysis. The velocity distribution of this radially vibrating polar cap can be expanded in terms of spherical harmonics as[13]

$$U_o(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}U_{lm}^{o}Y_{lm}(\theta,\phi), \tag{2}$$

where the spherical harmonics functions are defined as

$$Y_{lm}(\theta,\phi)=\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\,\theta)e^{jm\phi}, \tag{3}$$

$P_l^m(\cos\,\theta)$ is the *associated Legendre function* of degree l and order m evaluated at $\cos\,\theta$, and the spherical harmonic coefficients are given by

$$U_{lm}^{o}=\begin{cases}\dfrac{u_0}{2}\sqrt{\dfrac{4\pi}{2l+1}}[P_{l-1}^0(\cos\,\theta_s)-P_{l+1}^0(\cos\,\theta_s)], & m=0\\ 0, & m\neq 0.\end{cases} \tag{4}$$

For $l=0$, $P_{-1}^0(x)=1$.

After rotating the center of the vibrating polar cap to a new position with an elevation angle of θ_D and an azimuth angle of ϕ_D, the spherical harmonic coefficients of the obtained source velocity distribution on the sphere surface in the original coordinate can be written as[14–17]

$$U_{lm}^{D}=\sqrt{\frac{4\pi}{2l+1}}U_{l0}^{o}Y_{lm}^{*}(\theta_D,\phi_D), \tag{5}$$

where superscript "*" denotes the complex conjugate.

The center of the first source on the surface of Sphere 2 is assumed to be (θ_L,ϕ_L) in its own coordinate system O_2 and that of the second source on the surface of Sphere 3 be (θ_R,ϕ_R) in its coordinate system O_3; then the spherical harmonic coefficients of the velocity distribution of each source in its own coordinate systems are

$$U_{lm}^{L}=\sqrt{\frac{4\pi}{2l+1}}U_{l0}^{o}Y_{lm}^{*}(\theta_L,\phi_L), \tag{6}$$

$$U_{lm}^{R}=\sqrt{\frac{4\pi}{2l+1}}U_{l0}^{o}Y_{lm}^{*}(\theta_R,\phi_R). \tag{7}$$

The total sound field produced by the sources and the three spheres consists of three parts: the radiated and scattered sound fields from source 1 (Sphere 2) and source 2 (Sphere 3), and the scattered field from the listener's head (Sphere 1). It can be expressed in each coordinate system as

$$\begin{aligned}p_t(\mathbf{r}_{1r})&=p_{s_1}(\mathbf{r}_{1r})+p_{s_2}(\mathbf{r}_{2r})+p_{s_3}(\mathbf{r}_{3r}),\\ p_t(\mathbf{r}_{2r})&=p_{s_1}(\mathbf{r}_{1r})+p_{s_2}(\mathbf{r}_{2r})+p_{s_3}(\mathbf{r}_{3r}),\\ p_t(\mathbf{r}_{3r})&=p_{s_1}(\mathbf{r}_{1r})+p_{s_2}(\mathbf{r}_{2r})+p_{s_3}(\mathbf{r}_{3r}).\end{aligned} \tag{8}$$

The sound field (radiated and/or scattered) from each sphere can be conveniently expressed in its own coordinate system as

$$\begin{aligned}p_{s1}(\mathbf{r}_{1r})&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}C_{lm}h_l(kr_{1r})Y_{lm}(\theta_{1r},\phi_{1r}),\\ p_{s2}(\mathbf{r}_{2r})&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}D_{lm}h_l(kr_{2r})Y_{lm}(\theta_{2r},\phi_{2r}),\\ p_{s3}(\mathbf{r}_{3r})&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}E_{lm}h_l(kr_{3r})Y_{lm}(\theta_{3r},\phi_{3r}),\end{aligned} \tag{9}$$

where C_{lm}, D_m, and E_m are unknown spherical harmonic coefficients to be determined by applying the boundary conditions on the surfaces of the spheres. $k=\omega/c_0$ is the wave number and c_0 is the sound speed. $h_l(x)=j_l(x)-jn_l(x)$ is the *spherical Hankel function* of order l, $j_l(x)$ is the *spherical Bessel function* of order l, and $n_l(x)$ is the *spherical Neumann function* of order l.

The difficulty in solving the above equations is that there are functions and variables in three different coordinate systems. This can be simplified by using the translational addition theorem,[18–22] which expresses the sound pressure of the same point in space, originally represented on the coordinate system O_j in terms of the coordinate system O_i by

$$\begin{aligned}&h_l(kr_{jr})Y_{lm}(\theta_{jr},\phi_{jr})\\ &\quad=\sum_{p=0}^{\infty}\sum_{q=-p}^{p}Q_{pq}^{lm}(\mathbf{r}_{oij})j_p(kr_{ir})Y_{pq}(\theta_{ir},\phi_{ir}),\\ &\quad i,j=1,2,3,\end{aligned} \tag{10}$$

where $\mathbf{r}_{oij}=\mathbf{r}_{oj}-\mathbf{r}_{oi}=(r_{oij},\theta_{oij},\phi_{oij})$.

$$\begin{aligned}Q_{pq}^{lm}(\mathbf{r}_{oij})=&\sum_{n=|l-p|:2}^{l+p}4\pi(-j)^{p+n-l}h_n(kr_{oij})\\ &\times Y_{n,m-q}(\theta_{oij},\phi_{oij})g(m,l,-q,p,n),\end{aligned} \tag{11}$$

with $n=l+p,l+p-2,\ldots,|l-p|$, and

$$\begin{aligned}&g(m,l,-q,p,n)\\ &\quad=(-1)^m\sqrt{\frac{(2l+1)(2p+1)(2n+1)}{4\pi}}\\ &\qquad\times\begin{pmatrix}l & p & n\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}l & p & n\\ m & -q & -m+q\end{pmatrix}\end{aligned} \tag{12}$$

is related to the Gaunt coefficients, which are expressed as products of the Wigner $3j$ symbols, as shown in Eq. (12).[18,19,21,22]

Substituting Eqs. (9) and (10) into Eq. (8) and using the boundary condition $j\rho_0\omega v_n(\mathbf{r})=-\partial p_t(\mathbf{r})/\partial r$ at the surface of each rigid sphere in its own coordinate system (where ρ_0 is the density of the medium) yield

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lm}h_l'(ka_1)Y_{lm}(\theta_{1r},\phi_{1r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} D_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o12})j_p'(ka_1)Y_{pq}(\theta_{1r},\phi_{1r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o13})j_p'(ka_1)Y_{pq}(\theta_{1r},\phi_{1r}) = 0,$$

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o21})j_p'(ka_2)Y_{pq}(\theta_{2r},\phi_{2r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} D_{lm}h_l'(ka_2)Y_{lm}(\theta_{2r},\phi_{2r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o23})j_p'(ka_2)Y_{pq}(\theta_{2r},\phi_{2r}) = -j\rho_o c_o\sum_{l=0}^{\infty}\sum_{m=-l}^{l} U_{lm}^{L}Y_{lm}(\theta_{2r},\phi_{2r}),$$

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o31})j_p'(ka_3)Y_{pq}(\theta_{3r},\phi_{3r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} D_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o32})j_p'(ka_3)Y_{pq}(\theta_{3r},\phi_{3r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm}h_l'(ka_3)Y_{lm}(\theta_{3r},\phi_{3r}) = -j\rho_o c_o\sum_{l=0}^{\infty}\sum_{m=-l}^{l} U_{lm}^{R}Y_{lm}(\theta_{3r},\phi_{3r}). \quad (13)$$

The prime denotes the differentiation with respect to the argument of the function. Truncating the number of summations to L (depending on the calculation precision required) and equating the coefficients of $Y_{lm}(\theta_{1r},\phi_{1r})$, $Y_{lm}(\theta_{2r},\phi_{2r})$, and $Y_{lm}(\theta_{3r},\phi_{3r})$ in Eq. (13), the following coupled linear complex equations can be obtained:

$$C_{lm}h_l'(ka_1) + \sum_{p=0}^{L}\sum_{q=-p}^{p} D_{pq}Q_{lm}^{pq}(\mathbf{r}_{o12})j_l'(ka_1) + \sum_{p=0}^{L}\sum_{q=-p}^{p} E_{pq}Q_{lm}^{pq}(\mathbf{r}_{o13})j_l'(ka_1) = 0,$$

$$\sum_{p=0}^{L}\sum_{q=-p}^{p} C_{pq}Q_{lm}^{pq}(\mathbf{r}_{o21})j_l'(ka_2) + D_{lm}h_l'(ka_2) + \sum_{p=0}^{L}\sum_{q=-p}^{p} E_{pq}Q_{lm}^{pq}(\mathbf{r}_{o23})j_l'(ka_2) = -j\rho_o c_o U_{lm}^{L},$$

$$\sum_{p=0}^{L}\sum_{q=-p}^{p} C_{pq}Q_{lm}^{pq}(\mathbf{r}_{o31})j_l'(ka_3) + \sum_{p=0}^{L}\sum_{q=-p}^{p} D_{pq}Q_{lm}^{pq}(\mathbf{r}_{o32})j_l'(ka_3) + E_{lm}h_l'(ka_3) = -j\rho_o c_o U_{lm}^{R}. \quad (14)$$

For all l and $m(l=0,1,2,\ldots,L;m=-L,-L+1,\ldots,0,1,2,\ldots,L)$, the equations above can be expressed in a more compact form by using matrix notations as

$$\begin{bmatrix} \mathbf{S}_1 & \mathbf{Q}_{12} & \mathbf{Q}_{13} \\ \mathbf{Q}_{21} & \mathbf{S}_2 & \mathbf{Q}_{23} \\ \mathbf{Q}_{31} & \mathbf{Q}_{32} & \mathbf{S}_3 \end{bmatrix}\begin{bmatrix} \mathbf{C} \\ \mathbf{D} \\ \mathbf{E} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix}, \quad (15)$$

where the lmth element of the $(L+1)^2$ row vectors $\mathbf{A_1}$, $\mathbf{A_2}$, and $\mathbf{A_3}$ are 0, $-j\rho_o c_o U_{lm}^{L}$, and $-j\rho_o c_o U_{lm}^{R}$, respectively. The three groups of $(L+1)^2$ row unknown vectors are

$$\mathbf{C} = [C_{0,0} \quad C_{1,-1} \quad C_{1,0} \quad C_{1,1} \quad \cdots \quad C_{l,m} \quad \cdots \quad C_{L,L}]^T,$$

$$\mathbf{D} = [D_{0,0} \quad D_{1,-1} \quad D_{1,0} \quad D_{1,1} \quad \cdots \quad D_{l,m} \quad \cdots \quad D_{L,L}]^T,$$

$$\mathbf{E} = [E_{0,0} \quad E_{1,-1} \quad E_{1,0} \quad E_{1,1} \quad \cdots \quad E_{l,m} \quad \cdots \quad E_{L,L}]^T, \quad (16)$$

and the three groups of $(L+1)^2\times(L+1)^2$ diagonal complex matrices are

$$\mathbf{S}_1 = \mathrm{diag}\{s_{0,0}^{(1)} \quad s_{1,-1}^{(1)} \quad s_{1,0}^{(1)} \quad s_{1,1}^{(1)} \quad \cdots \quad s_{l,m}^{(1)} \quad \cdots \quad s_{L,L}^{(1)}\},$$

$$s_{l,m}^{(1)} = h_l'(ka_1), \quad \forall\, m,$$

$$\mathbf{S}_2 = \mathrm{diag}\{s_{0,0}^{(2)} \quad s_{1,-1}^{(2)} \quad s_{1,0}^{(2)} \quad s_{1,1}^{(2)} \quad \cdots \quad s_{l,m}^{(2)} \quad \cdots \quad s_{L,L}^{(2)}\},$$

$$s_{l,m}^{(2)} = h_l'(ka_2), \quad \forall\, m,$$

$$\mathbf{S}_3 = \mathrm{diag}\{s_{0,0}^{(3)} \quad s_{1,-1}^{(3)} \quad s_{1,0}^{(3)} \quad \cdots \quad s_{1,1}^{(3)} \quad \cdots \quad s_{L,L}^{(3)}\},$$

$$s_{l,m}^{(3)} = h_l'(ka_3), \quad \forall\, m \quad (17)$$

The element of $(L+1)^2\times(L+1)^2$ complex matrix $\mathbf{Q}_{ij}$ is

$$\xi_{lm,pq}^{ij} = Q_{lm}^{pq}(\mathbf{r}_{oij})j_l'(ka_i), \quad (18)$$

where the row index is lm and the column index is pq, in the orders of

$$lm = \{(0,0)(1,-1)(1,0)(1,1)\cdots(l,m)\cdots(L,L)\},$$

$$pq = \{(0,0)(1,-1)(1,0)(1,1)\cdots(l,m)\cdots(L,L)\}. \quad (19)$$

Having obtained the $3(L+1)^2$ unknown spherical harmonic coefficients C_{lm}, D_{lm}, and E_{lm}, the total sound field can be calculated by using any equation in Eq. (8). For example,

$$p_t(\mathbf{r}_{1r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} C_{lm} h_l(kr_{1r}) Y_{lm}(\theta_{1r}, \phi_{1r}) + \sum_{l=0}^{L}\sum_{m=-l}^{l} D_{lm} h_l(kr_{2r}) Y_{lm}(\theta_{2r}, \phi_{2r}) + \sum_{l=0}^{L}\sum_{m=-l}^{l} E_{lm} h_l(kr_{3r}) Y_{lm}(\theta_{3r}, \phi_{3r}). \quad (20)$$

It is often convenient to designate the complex acoustic pressure p as the output and the complex source volume acceleration $j\omega\rho_0 q/4\pi$ as the input.[2] For the radially vibrating cap source, $q=2\pi a_i^2(1-\cos\theta_s)u_0$. The relation between the source input signal and the receiving output signal or the frequency response from one sound source to the receiving point can be obtained the same way as that for Eq. (20). Nevertheless, the given source spherical harmonic coefficients used now are the followings instead of Eqs. (6) and (7),

$$U_{lm}^{L} = \sqrt{\frac{4\pi}{2l+1}} \frac{U_{l0}^{o} Y_{lm}^{*}(\theta_L, \phi_L)}{j\omega\rho_0 a_2^2(1-\cos\theta_s)u_0/2}. \quad (21)$$

$$U_{lm}^{R} = \sqrt{\frac{4\pi}{2l+1}} \frac{U_{l0}^{o} Y_{lm}^{*}(\theta_R, \phi_R)}{j\omega\rho_0 a_3^2(1-\cos\theta_s)u_0/2}. \quad (22)$$

Obviously, this model can be used to model similar systems where the radii of the spheres are different or when the vibrating cap sound sources are located at different positions on the sphere with different apertures. The model can also be extended to more sophisticated settings with multiple listeners and loudspeakers.[23]

III. NUMERICAL RESULTS AND DISCUSSION

A. Sound source characteristics

The sound sources used in the simulations include a single point source, a baffled piston, a single radially vibrating cap on a rigid sphere, and a radially vibrating cap on a rigid sphere with another rigid sphere aside. The single receiving point is on the central axis line of the sound source with a distance of 1.0 m from the source acoustic center. If the source is a vibrating cap on the surface of a sphere, the distance is that between the center of the cap on the surface of the sphere (not the center of the sphere) and the receiving point. The radius of the baffled piston is 0.1 m and the radius of source spheres is $a_1=a_2=0.1$ m. Half of the cap source span angle, θ_s, is 30°, which is about the size of a 4 in. loudspeaker. For settings with two source cases where another sphere is introduced, the two sources subtend angles of 60° and 20° at the receiving position, and the distance between the centers of the two spheres are 1.1 and 0.382 m, respectively. Even though the "stereo dipole" configurations[2] require the two sources to be closely spaced with only a 10° subtended angle, the size of the source sphere in this paper is too large for a 10° subtend angle, so that a subtended angle of 20° is used in the simulations. The locations of the vibrating cap center on the surface of the left sphere (Sphere 2) are ($\theta_L=150°$, $\phi_L=0°$) and ($\theta_L=170°$, $\phi_L=0°$) in coordinate system O_2, respectively, for the 20° and 60° subtended angle configurations, and the locations of the vibrating cap on the surface of right sphere (Sphere 3) are ($\theta_R=150°$, $\phi_R=180°$) and ($\theta_R=170°$, $\phi_R=180°$) in coordinate system O_3.

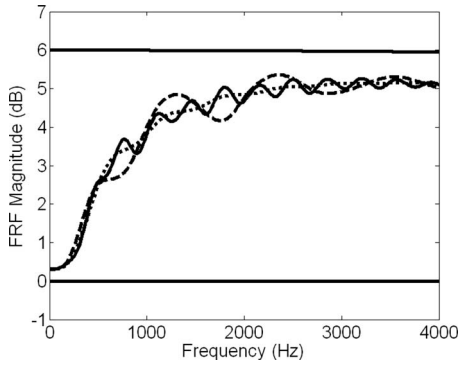

FIG. 2. Magnitude of the calculated FRFs for various sound source configurations. The lower solid line with 0 dB magnitude is that of the point source, the upper solid line with a magnitude of 6 dB is that of a baffled piston, the dotted curve is that of a single cap source on a rigid sphere, and the middle solid and dashed curves are for the two sphere sound sources with 60° and 20° subtended angles, respectively.

In a series of preliminary calculations, it was detected that using a spherical harmonics expansion degree of $L=10$ provides sufficient iteration precision, with the errors smaller than 1% for the current setup in the frequency range up to 4 kHz. Higher expansion degrees should be used for larger ratio of the size of the sphere to the wavelength. As this acoustical model can only approximate the characteristics of a loudspeaker up to a certain frequency range with the wavelength larger than the source size, no attempt has been made to obtain results at higher frequencies. Nevertheless, the range up to 4 kHz covers the basic and most important frequency range in fields such as communication and sound reproduction.

The frequency response functions (FRFs) are calculated for all above mentioned configurations with a source input of the same volume acceleration at 257 evenly distributed points between 0 and 4096 Hz, and the results are presented in Fig. 2. The lower solid line with 0 dB magnitude is the FRF of the point source, and the upper solid line with a magnitude of 6 dB is that of the baffled piston. The dotted curve is the FRF of a single cap source on a rigid sphere, and finally the middle solid and dashed curves are that of the two sphere cases. The second source on the right sphere is not active and the sphere is assumed to be rigid for the configurations with two sound sources, so that the frequency response is just that of one source. As expected, the sources of a vibrating cap on a rigid sphere do not have a flat frequency response (in comparison to an ideal point source). The vibrating cap on a rigid sphere radiates like a point source in the low frequency range. In the high frequency range, it almost behaves like a baffled piston and the pressure it radiates is nearly two times higher, which is caused by the scattering of the rigid sphere where the cap sound source is located.

From the FRF curves of the configurations with two rigid sphere sound sources, it can be seen that the scattering from a neighboring rigid sphere adds extra fluctuations on the original FRF curves. The magnitude and frequency of the

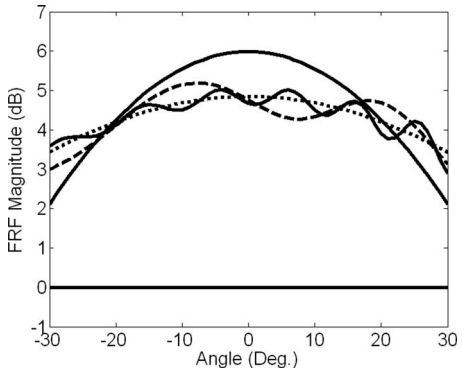

FIG. 3. Magnitude of the calculated FRFs at 2000 Hz when the receiving point swings from −30° (left) to 30° (right) for the configurations of Fig. 2. The distance between the source and the receiving point remains at 1.0 m. The lower solid line is that for the point source, the upper solid curve is that for the baffled piston, the dotted curve is that for the single cap source on a rigid sphere, and the middle solid and dashed curves are that for the two sphere sound sources with 60° and 20° subtended angles, respectively.

fluctuations depend on the distance between the two spheres, and the fluctuation magnitude becomes larger for nearer distance. The magnitude variation, for example, caused by the neighboring rigid sphere scattering can be as large as about 1 dB around 2000 Hz. It also can be observed that the magnitude of the fluctuation decreases at higher frequencies. This is because the cap source behaves more directional at higher frequencies, so less sound is radiated to and scattered from the neighboring sphere.

Figure 3 illustrates the directional behavior of the simulated systems, which depicts the calculated transfer functions at a specific frequency of 2000 Hz. At this frequency, the receiving point swings from −30° (left) to 30° (right) (with the distance between the source and the receiving point remaining at 1.0 m). 0° is the angle for the central axis line. The lower solid line of 0 dB is for the monopole source. The upper solid curve is that of the baffled position, the dotted curve is for one rigid sphere sound source, and the middle solid and dashed curves are for the two rigid sphere sound sources that subtend angles of 60° and 20°, respectively. As expected, the vibrating cap on a rigid sphere presents a directional behavior and the magnitude of the FRF at 30° can be nearly 2 dB lower than that at 0°. The scattering from the neighboring rigid sphere makes the sound pressure spatial distribution more complicated. This might result in a large difference in the excess sound pressure level between two ears of a CTC system.

B. CTC systems with non-omnidirectional sound sources

Assume that $\mathbf{C}$ is the 2×2 plant transfer function matrix in a CTC system, $\mathbf{V}=[v_1,v_2]^T$ is the source input signal vector, $\mathbf{H}$ is the 2×2 CTC matrix being designed, and $\mathbf{U}=[u_1,u_2]^T$ is the recorded/synthesized binaural signal vector. The system output $\mathbf{W}=[w_1,w_2]^T$ can be expressed as

$$\mathbf{W}=\mathbf{CV}=\mathbf{CHU}. \tag{23}$$

In order to ensure the reproduction of the binaural signals at the receiving points, $\mathbf{W}=\mathbf{U}e^{-j\omega\Delta}$, where Δ is a delay to ensure causality, it is necessary that $\mathbf{CH}=\mathbf{I}e^{-j\omega\Delta}$; thus $\mathbf{H}=\mathbf{C}^{-1}e^{-j\omega\Delta}$. However, the matrix could be ill-conditioned in some situations, so the following approximate solution is often used:[2–5]

$$\mathbf{H}=[\hat{\mathbf{C}}^{\mathrm{H}}\hat{\mathbf{C}}+\beta\mathbf{I}]^{-1}\hat{\mathbf{C}}^{\mathrm{H}}e^{-j\omega\Delta}, \tag{24}$$

where $\hat{\mathbf{C}}$ is the estimation of $\mathbf{C}$, and for the matched case, $\hat{\mathbf{C}}=\mathbf{C}$. In this paper, $\hat{\mathbf{C}}$ is also referred to as the setup plant transfer functions or setup HRTFs, while $\mathbf{C}$ is referred to as the playback plant transfer functions or HRTFs. β is a regularization parameter which constraints the energy of the source input signals and prevents the singularity of $\hat{\mathbf{C}}^{\mathrm{H}}\hat{\mathbf{C}}$ from saturating the filter gains. The performance of a CTC system is described by the performance matrix below:

$$\mathbf{P}=\mathbf{CH}, \tag{25}$$

where $\mathbf{P}$ should be a unit matrix under ideal CTC and the elements of the matrix should be $P_{11}=P_{22}=1$ and $P_{12}=P_{21}=0$. Channel separation is used here to illustrate the performance variation with frequency, which is defined as the ratio of the ipsilateral ear response to the contralateral ear response from the left (or right) recorded/synthesized binaural signal source:[4,5,24]

$$J_L(\omega)=20\log_{10}(|P_{11}(\omega)/P_{21}(\omega)|),$$

$$J_R(\omega)=20\log_{10}(|P_{22}(\omega)/P_{12}(\omega)|). \tag{26}$$

For the left-right symmetrical configurations, the right to left channel separation J_R (for the right recorded/synthesized signal) and the left to right channel separation J_L (for the left recorded/synthesized signal) are equal. The channel separations are thus interchangeably and the term "channel separation" is used to refer to both.

The commonly used models for studies dealing with CTC systems are the free field model and the spherical head model, which both assume that the playback sound sources are ideal point sources. The model shown in Fig. 1 provides the basis for this paper and is used to study non-ideal sound sources and their interaction. Here, the distance between the source center and the center of the listener head is 1.0 m, and the two sound sources subtend an angle of 60° at the listener position, the radius of source spheres is $a_1=a_2=0.1$ m, and the radius of receiver sphere (head) is $a_3=0.09$ m.

To emphasize the effects of the sound source directivity on the CTC performance, the spherical head model with two ideal point sources is used first of all by assigning different gains at different directions on its playback plant transfer functions. More complicated situations involving orientation changes of the playback sound sources are studied in Secs. III C and III D. Figure 4 shows the simulation results of the corresponding CTC systems. The dotted line corresponds to the ideal omnidirectional playback sound sources and the solid lines correspond to the playback sound sources without omnidirectional directivities. Note that a channel separation greater than 50 dB is clipped at 50 dB throughout the paper.

In the above simulations, the setup plant transfer functions obtained from the spherical head model with two ideal point sources are used to design the CTC filters. While cre-

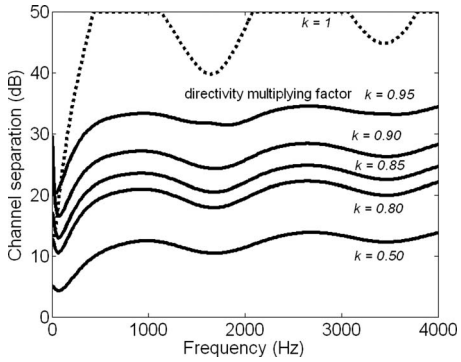

FIG. 4. Channel separations of the CTC system using playback plant transfer functions with different directivities. The setup plant transfer functions are obtained with the point sources of the spherical head model. The playback sound sources are the same as the setup ones (dotted line) or the ones that radiate less in the crosstalk path direction with the transfer function magnitude k (k=0.95, 0.90, 0.85, 0.80, and 0.50) times of that of the original (solid curves, the curve with the highest channel separation of above 30 dB is that for 0.95, then 0.90, 0.85, 0.80, and the lowest curve is that for 0.50).

ating the filters, the regularization parameter β is adjusted to make the channel separation just greater than 40 dB above 500 Hz under the matched situation throughout the paper. In this simulation, β is 0.012. The directivities of the playback plant transfer function are simulated by directly multiplying small values of 0.95, 0.90, 0.85 0.80, and 0.5 on the crosstalk path transfer functions from the same spherical head model with two ideal point sources, which corresponds to differences between the signals at two ears of about 0.4, 0.9, 1.4, 1.9, and 6 dB of excess sound pressure level. This simple approximation might be too rough for modeling the frequency-dependent directivities of real loudspeakers, yet it shows clearly that a small difference in the magnitude between two responses at two ears reduce the channel separation significantly. For example, it can be seen from the figure that if the excess sound pressure level difference between two ears caused by the source directivity is about 2 dB, and the channel separation drops significantly from above 40 dB to less than 20 dB.

C. Dynamic CTC systems with non-ideal sound sources

Figure 5 shows an example of a dynamic CTC system where a listener moves to a location which is not at the center of a symmetrical loudspeaker setup.[11] This situation

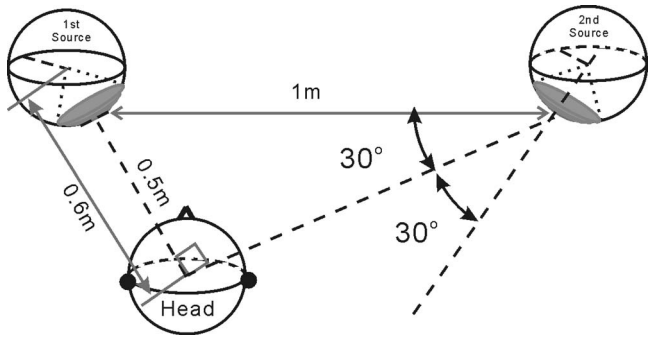

FIG. 5. A dynamic CTC system with the listener being located at an asymmetrical position.

can often be met in a CAVE-like environment where the user can walk around freely inside. At the stage of designing the binaural virtual sound imaging system for a CAVE, the plant transfer functions (or HRTFs) from a database are used as the setup ones to design a set of CTC filter for each potential listener position and direction in the CAVE. It should be emphasized that these plant transfer functions (or HRTFs) in the database were usually measured in prior with an ideal omnidirectional sound source for sound from different distances and directions. However, in practical applications in a CAVE, the loudspeaker is rarely ideal and usually radiates differently in different directions, and it might also not be realistic sometimes for the loudspeakers to be rotated in realtime according to the listener's position.

In the example shown in Fig. 5, the setup plant transfer functions are obtained with ideal point sources, the playback sound sources are the vibrating cap sources, their orientations were fixed in prior, and the listener at the present position is not on the central axis line of the right source. The distance between the two source centers is 1.0 m and the distances from the listener's head to the left and the right sources are 0.5 and 0.866 m, respectively. In the coordinate system O_1, the center of the listener head is $\mathbf{r}_{o1}=(0,0,0)$, the center of the left source sphere is $\mathbf{r}_{o2}=(0.6,30°,180°)$, and that of the right source sphere is $\mathbf{r}_{o3}=\mathbf{r}_{o2}+(1.1,90°,0°)$. The location of the vibrating cap center on the surface of the left sphere is (θ_L=150°, ϕ_L=0°) in its own coordinate system O_2, and the location of the vibrating cap on the surface of the right sphere is (θ_R=150°, ϕ_R=180°) in its own coordinate system O_3.

The left to right and the right to left channel separations of the CTC system, which are based on different playback plant transfer functions, are shown in Fig. 6. The regularization parameter β is 0.035. The dotted curves show the channel separation when the playback sound sources are the same as the setup ideal point sources, while the solid curves are for vibrating cap playback sources in the original pre-fixed orientation. The dashed curves show the channel separation, while the orientation of the right source of Fig. 5 is adaptively rotated to face the listener so that the listener is on the central axis line of the sound sources. This kind of rotation is sometimes hard to be implemented in practice; however, its simulation results are included in the figure to show the effects of the source orientation on the channel separation.

For this configuration, the center of the right source sphere is at $\mathbf{r}_{o3}=(0.966,60°,0°)$ and the location of the vibrating cap on its surface is (θ_R=120°, ϕ_R=180°) in its own coordinate system O_3. By comparing the dashed and dotted curves, it can be observed that the mismatch between the spatial and spectrum characteristics of the point sources and cap sources reduces the channel separation from over 40 dB to around 30 dB. By comparing the dashed and solid curves, it can be seen that the channel separation will be further reduced by a few decibels for the pre-fixed source orientations, especially in higher frequency range due to the higher directivity of the cap source. Unfortunately, it is often hard in practice to have the orientation of the right source rotated in realtime to face the listener so that the listener is on the original pre-fixed direction or central axis line of it. It can

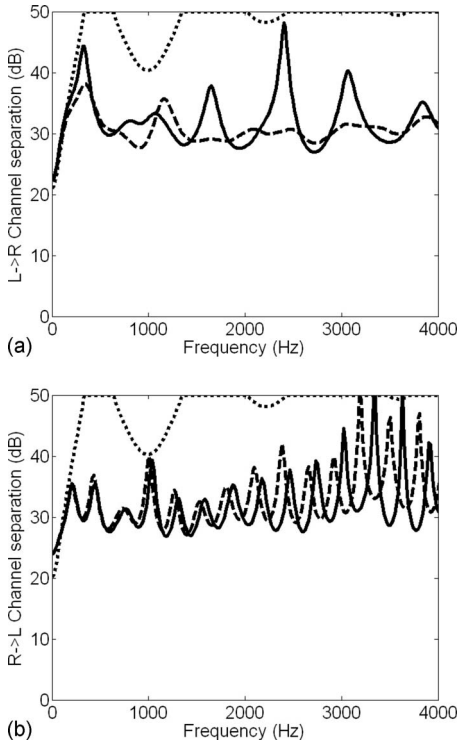

FIG. 6. Channel separations of a dynamic CTC system using different playback sound sources. The setup plant transfer functions are obtained with the point sources instead of the vibrating cap sources in Fig. 5. The playback sound sources are the same as the setup ones (dotted curve), or are the vibrating cap sources with the right source being rotated to face the listener head (dashed curve), or are the vibrating cap sources in original pre-fixed orientation (solid curve). (a) The left to right channel separation. (b) The right to left channel separation.

also be found that the right to left channel separation fluctuates more frequently than the left to right one, and this is because the distance between the listener's head and the right source sphere is larger than that with the left one.

The magnitude difference of the frequency responses between the present and the original pre-fixed orientations can be larger than 3 dB at 4000 Hz due to the directional properties of the source. Nevertheless, the performance loss is much smaller than that shown in Fig. 4, which illustrates that a small alteration of only 0.9 dB in terms of the difference in excess sound pressure level between two ears caused by the source directivity can reduce the channel separation from more than 40 to about 25 dB. The smaller performance loss for the dynamic CTC system can be explained by the fact, that even though the magnitude difference of the frequency responses in the present and the original pre-fixed directions caused by the source directional properties is larger than 3 dB at 4000 Hz, the change in the excess sound pressure level difference between two ears caused by the change in source orientation is much smaller than 3 dB as the two ear subtended angle seen from the source is much smaller than the value of the change in the source orientation angle (30° here). With the increase in the distance between the listener and the source, the two ears' subtended angle seen from the source becomes narrower, and this weakens the influence of the source directional properties on the channel separation of the dynamic CTC systems.

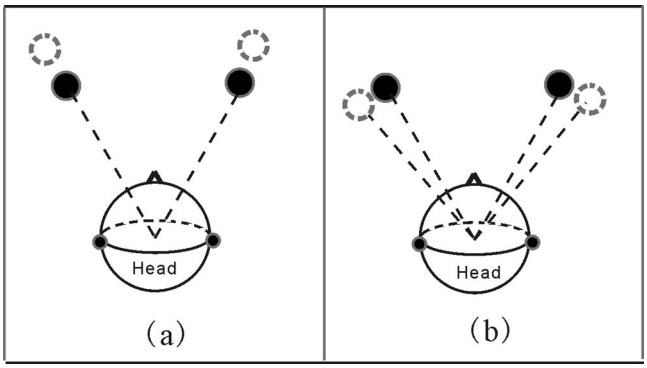

FIG. 7. The configurations for the CTC system with misaligned ideal point sound sources: (a) sources and the listener distance misalignment; (b) sources' subtended angle misalignment.

D. CTC systems with misaligned sound sources

In order to highlight the basic mechanisms, the impact of misalignments of the source-listener distance and the source subtended angle are illustrated by using the spherical head model with two ideal point sources, as shown in Fig. 7, while the effect of source orientation misalignment is illustrated by the cap sources. The channel separations of the CTC system with playback sound sources at different distance are depicted in Fig. 8, where the setup plant transfer functions are obtained when the distance between each of the sources and the listener is 1.0 m and the two sound sources subtend an angle of 60°. The playback functions are the same as the setup functions or obtained by only changing the distance from each of the sources to the listener to 1.1, 1.2, 1.3, 1.4, and 1.5 m along the original source-listener lines, as shown in Fig. 7(a). It can be observed from the figure that a small change in the distance along the original source-listener lines does not have a significant impact on the channel separations. Even if the distance is increased by 50% to 1.5 m, the average channel separation is still greater than 30 dB.

Figure 9 shows the channel separations of the CTC sys-

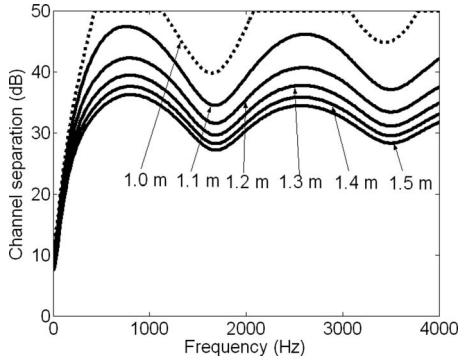

FIG. 8. Channel separations of the CTC system with distance misaligned playback sound sources. The setup plant transfer functions are obtained with the spherical head model with two ideal point sources when the distance between each of the sources and the listener is 1.0 m and the two sound sources subtend an angle of 60°. The playback ones are the same as the setup ones (dotted curve) or obtained by changing the distance from each of the sources to the listener to 1.1, 1.2, 1.3, 1.4, and 1.5 m (solid curves, the curve with the highest channel separation is that for 1.1 m, then 1.2 m, 1.3 m, 1.4 m, and the lowest curve is that for 1.5 m).

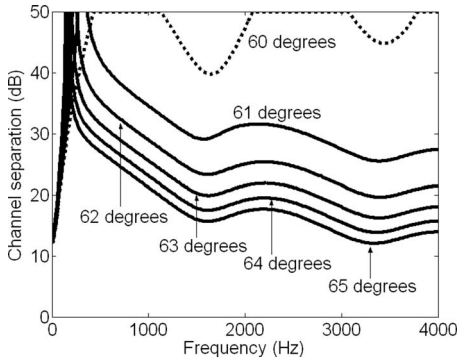

FIG. 9. Channel separations of the CTC system with subtended angle misaligned playback sound sources. The setup plant transfer functions are the same as that in Fig. 7. The playback ones are the same as the setup ones (dotted curve) or obtained by increasing the two source subtended angle from 60° with 1°, 2°, 3°, 4°, and 5° (solid curves, the curve with the highest channel separation is that for 1°, then 2°, 3°, 4°, and the lowest curve is that for 5°).

tem with the playback sound sources at different subtended angles as the configurations shown in Fig. 7(b). The setup plant transfer functions are the same as that used for Fig. 8. The playback functions are the same as the setup functions or obtained by increasing the two source subtended angle from 60° by 1°, 2°, 3°, 4°, and 5°, while the distance from the source to listener remains at 1.0 m. The figure underlines that changing the sources' subtended angle has a larger influence on the channel separation than changing the distance. The average channel separation drops to nearly 30 dB for a 1° change and to nearly 15 dB for a 5° change. It appears that this kind of change causes larger differences between the transfer functions between two ears and the source than those caused by changing the distance.

The loudspeaker box center (for example, the rigid sphere center in the model instead of the cap center) is sometimes improperly treated as the acoustic center of the sound source in practical applications. If the source orientation misalignment is referred to this box center, the channel separation drops significantly from more than 40 dB to about 20 dB with a misalignment of only 10°. Detailed analyses of preliminary simulations show that the main reason for this performance loss is the change in the sound source subtended angle seen from the listener, which is caused by the rotation of the sound source whose physical center (box or rigid sphere center) is not the acoustic center.[25] The performance loss caused by the misalignment of the subtended angle of the sources has already been shown in Fig. 9. In the following simulations, as shown in the configurations of Fig. 10, the relative positions (angle and the distance) between the acoustic center of the sound sources (the center of the sound cap) and the listener remain unchanged and only the orientations of the cap sources are changed to focus on the influence of the directional properties of the source.

Before the orientation of the cap source changes, the center of the listener head is $\mathbf{r}_{o1}=(0,0,0)$, the center of the left source sphere is $\mathbf{r}_{o2}=(1.1,30°,180°)$, and that of the right source sphere is $\mathbf{r}_{o3}=(1.1,30°,0°)$ in the coordinate system O_1 of Fig. 1. The location of the vibrating cap center on the surface of the left sphere is $(\theta_L=150°,\ \phi_L=0°)$ in its own coordinate system O_2 and that of the right vibrating cap is $(\theta_R=150°,\ \phi_R=180°)$ in its own coordinate system O_3. After rotating the orientation of the cap source α degree outside (positive value) or inside (negative value) with the center of cap source, the center of the left source sphere is $\mathbf{r}_{o2}=(1.0,30°,180°)+(0.1,30°-\alpha,180°)$ and that of the right source sphere is $\mathbf{r}_{o3}=(1.0,30°,0°)+(0.1,30°-\alpha,0°)$. The location of the vibrating cap center on the surface of the left sphere is $(\theta_L=150°+\alpha,\ \phi_L=0°)$ in its own coordinate system O_2 and the location of the vibrating cap on the surface of right sphere is $(\theta_R=150°+\alpha,\ \phi_R=180°)$ in its own coordinate system O_3. In the simulation results shown here, α was set as 15°.

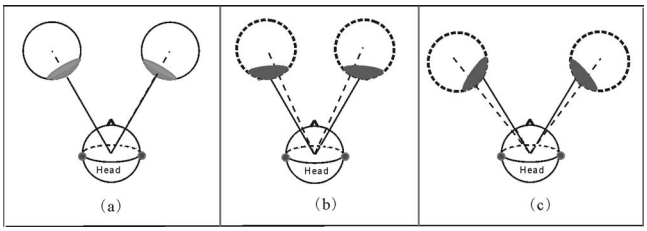

FIG. 10. The configurations for the CTC system with orientation misaligned directional sound sources. (a) The listener is on the central axis line of the sound sources, (b) the orientation of the cap sources is rotated 15° outside with the center of cap source, and (c) the orientation of the cap sources is rotated 15° inside with the center of cap source.

The channel separations of the CTC system caused by a change in the orientation of the directional source are shown in Fig. 11 where the regularization parameter β is 0.03. It becomes evident that the channel separation decreases significantly from more than 40 dB to about 25 dB, and the channel separation is larger for higher frequency because of the stronger directional properties of the source at high frequencies. It is also interesting to note that there are some frequencies (peaks in the curves) where the channel separation does not drop so much. These frequencies correspond to the impact of the scattering from the neighboring sphere, where the frequency responses differences at the two ears that can be observed for different configurations might be similar due to the scattering. By comparing the dashed curve and the solid curve, it can be found that the channel separa-

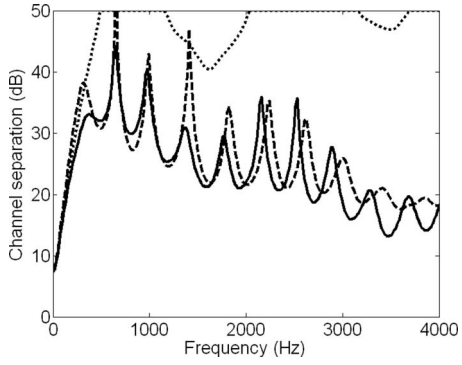

FIG. 11. Channel separations of the CTC system caused by the orientation change in the directional sources for the 60° source span system. The setup plant transfer functions are obtained with the original orientation. The playback ones are the same as the setup ones (dotted curve) or obtained by rotating the orientation of the cap sources 15° outside (dashed curve) or inside (solid curve) with the center of cap source.

tion is lower when the playback sound sources are rotated inside (solid curve). This might be explained by the larger scattering from the neighboring sphere due to the directivity of the cap source at high frequencies when the playback sound sources are rotated inside.

IV. DISCUSSIONS

By comparing the results of Fig. 11 with Fig. 9, where the two sources subtended angle changes by only a few degrees, it can be concluded that the influence on the channel separation of the CTC system, caused by the change in the directional source orientation, is smaller than the influence of the misalignment of the two sources subtended angle. Similar simulations were also carried out for the 20° source span system and when the source orientation rotates to different degrees. As the results are similar, they are not presented in this paper. All these results show that a slight change in the distance between the sources and the listener (still on the source-listener axis line) causes only a minor performance loss. The directional properties of the playback sound sources are responsible for a slight performance loss, especially at high frequencies, while the scattering from the neighboring sphere sometimes increases and sometimes decreases the channel separation in the middle frequency range. However, the most sensitive source misalignment is triggered by changing the subtended angle of the playback sound sources seen from the listener.

Extensive experiments have been carried out in the semi-anechoic chamber at the Institute of Technical Acoustics of the RWTH Aachen University to investigate the effects of using different playback sound sources, the channel separations of the dynamic CTC system with different source orientations, and the channel separations of the CTC system with misaligned sound sources.[26] The trends that could be detected from those experiment results are in concordance with that of the numerical simulations presented in this paper, so they are not given in this paper for conciseness. These results clearly show that the playback sound sources also play an important role in CTCs. Even though the different types of sources, the spatial characteristics (or directivity in the far field) of the sources, and positions of the sound sources all have a certain influence on the performance of the system, the channel separation of the CTC is most sensitive to the subtended angle of the playback sound sources. If highly accurate binaural cues are required in practical applications, the type and characteristics of the playback sound sources, their locations, and orientations all should be considered carefully.

For a CTC system, the factors that make the playback HRTFs mismatch the setup ones include the variation in individual HRTFs, the misalignment of the listener's head position and orientation, the misalignment of position and orientation of playback sound sources, the variation in the spatial and temporal characteristics of the playback sound sources, and the variation in the playback acoustics environment. The channel separation can decrease more than 15 dB just due to the difference between two individuals' HRTFs, and that caused by the position and orientation misalignment of the listener's head and/or the playback sound sources can be even larger depending on the degree of misalignment. It has also been found that the variations in the playback sound sources and the playback acoustics environment both can decrease the performance of the CTC system significantly. From all results obtained in this paper and the earlier studies by others on the robustness of the CTC systems, it can be deduced that even though any kind of playback and setup transfer function mismatches can reduce the channel separation performance of the CTC system to a certain extent, the CTC system is more sensitive to the mismatch that makes the change in propagation time from a source to one ear be in opposite direction to that to the other ear (reverse time response structures at two ears).[4–6,11]

For binaural virtual sound imaging systems that use two loudspeakers together with the CTC techniques, it has been proved that the system can only deliver accurate binaural cues when the setup and playback plant transfer functions are closely matched.[3] Two kinds of approaches can be used to address this problem: the dynamic (adaptive) CTC system and the robust design. The robust design means to optimize the physical properties (for example, locations of the sound sources) of the CTC systems so that its performance is not sensitive to the mismatch.[4,5] To implement either approach, the relation between the performance change and the basic parameters of the system need to be understood. From previous research and the findings in this paper, it is clear that a small change in the magnitude and phase difference between the transfer functions of two ears might reduce the performance of a CTC system significantly because of the coherent canceling between two waves. In the high frequency range, a small misalignment in space can cause a large change in phase, so that the robustness of the CTC system decreases in the high frequency range. Fortunately, human heads and loudspeaker boxes usually show more directionality in the high frequency range, and a hybrid system that uses both CTC and the inherent natural channel separation capabilities of sound sources and human head might offer more robustness and better channel separation performance.[4,24]

V. CONCLUSIONS

In the field of virtual acoustic imaging studies, there are several analytical models that are usually used. In the models such as the free field model and the spherical head model, the sound sources involved are treated as omnidirectional point sources. Based on spherical harmonics decomposition and the translational addition theorem, an analytical model has been developed in this paper that can take the scattering and spatial characteristics of the sound sources into consideration. The model is universal and could be useful when it comes to modeling similar systems with different sizes of spheres and with different positions and apertures of the cap sound sources. The model can also be extended and thus used for more sophisticated applications with multiple listeners and loudspeakers.

The effects of the spatial characteristics of a sound source and a misalignment on the performance of the CTC systems were studied based on the proposed model. It was

found that despite the fact that the different types of sources, the spatial characteristics of the sources, and the positions of the sound sources all have a certain influence on the system performance, the channel separation of the CTC is most sensitive to the misalignment of the subtended angle of the playback sound sources. The influence of the spatial characteristics of the source on the channel separation can be diminished somehow by placing the sources as far away from the listener as possible, so that the subtended angle from the source to two ears of the listener becomes smaller.

Future work can be extending the proposed model to investigate the influence of large reflective walls on the CTC system quantitatively and to investigate more sophisticated settings with multiple listeners and loudspeakers in a CAVE-like environment or an ordinary room.

ACKNOWLEDGMENTS

X.Q. is a research fellow of the Alexander von Humboldt Foundation and this research project is partially supported by Project No. 10674068 of NSFC. B.M. is a scholarship holder from CNPq, Brazil.

[1]M. Vorländer, *Auralization* (Springer-Verlag, Berlin, 2008), pp. 279–298.
[2]P. A. Nelson, M. Park, T. Takeuchi, and F. M. Fazi, "Binaural hearing and systems for sound reproduction (A)," J. Acoust. Soc. Am. **123**, 3469 (2008); in Proceedings of Acoustics'08 Conference, Paris, France (2008), pp. 3531–3536.
[3]M. A. Akeroyd, J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," J. Acoust. Soc. Am. **121**, 1056–1069 (2007).
[4]M. R. Bai and C. Lee, "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," J. Acoust. Soc. Am. **120**, 1976–1989 (2006).
[5]T. Takeuchi, P. A. Nelson, and H. Hamada, "Robustness to head misalignment of virtual sound imaging systems," J. Acoust. Soc. Am. **109**, 958–971 (2001).
[6]J. Rose, P. A. Nelson, B. Rafaely, and T. Takeuchi, "Sweet spot size of virtual sound imaging systems at asymmetric listener locations," J. Acoust. Soc. Am. **112**, 1992–2002 (2002).
[7]J. Daniel, R. Nicol, and S. Moreau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," in AES 114th Convention (2003), Paper No. 5788.
[8]A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am. **93**, 2764–2778 (1993).
[9]P. Gauthier and A. Berry, "Adaptive wave field synthesis for active sound field reproduction: Experimental results," J. Acoust. Soc. Am. **123**, 1991–2002 (2008).
[10]D. B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment," J. Acoust. Soc. Am. **110**, 1195–1198 (2001).
[11]T. Lentz, "Binaural technology for virtual reality," Ph.D. thesis, RWTH Aachen University, Aachen, Germany (2007).
[12]J. Garcia-Bonito and S. J. Elliott, "Local active control of diffracted diffuse sound fields," J. Acoust. Soc. Am. **98**, 1017–1024 (1995).
[13]Z. Lin, J. Lu, C. Shen, X. Qiu, and B. Xu, "Active control of radiation from a piston set in a rigid sphere," J. Acoust. Soc. Am. **115**, 2954–2963 (2004).
[14]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999), pp. 183–232.
[15]M. Pollow, "Variable directivity of dodecahedron loudspeakers," Ph.D. thesis, RWTH Aachen University, Aachen, Germany (2007).
[16]D. Healy, D. Rockmore, P. Kostelec, and S. Moore, "FFTs for the 2-sphere-improvements and variations," J. Fourier Anal. Appl. **9**, 341–385 (2003).
[17]F. Zotter and R. Höldrich, "Modeling radiation synthesis with spherical loudspeaker arrays," in Proceedings of 19th International Congress on Acoustics, Madrid, Spain (2007).
[18]W. C. Chew, "Recurrence relations for three-dimensional scalar addition theorem," J. Electromagn. Waves Appl. **6**, 133–142 (1992).
[19]G. C. Gaunaurd, H. Huang, and H. C. Strifors, "Acoustic scattering by a pair of spheres," J. Acoust. Soc. Am. **98**, 495–507 (1995).
[20]H. Huang and G. C. Gaunaurd, "Acoustic scattering of a plane wave by two spherical elastic shells above the coincidence frequency," J. Acoust. Soc. Am. **101**, 2659–2668 (1997).
[21]N. A. Gumerov and R. Duraiswami, "Computation of scattering from N spheres using multipole reexpansion," J. Acoust. Soc. Am. **112**, 2688–2701 (2002).
[22]K. M. Li, W. K. Lui, and G. H. Frommer, "The diffraction of sound by an impedance sphere in the vicinity of a ground surface," J. Acoust. Soc. Am. **115**, 42–56 (2004).
[23]B. Masiero and X. Qiu, "Two listeners crosstalk cancellation system modelled by four point sources and two rigid spheres," Acta. Acust. Acust. **95**, 379–385 (2009).
[24]W. G. Gardner, "3-D audio using loudspeakers," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, (1997).
[25]X. Qiu, M. Vorländer, and B. Masiero, "Effects of practical loudspeaker characteristics on virtual acoustic imaging systems (A)," J. Acoust. Soc. Am. **123**, 3202 (2008); in Proceedings of Acoustics'08 Conference, Paris, France (2008), pp. 1719–1724.
[26]X. Qiu, B. Masiero, and M. Vorländer, "Experimental study on channel separation of crosstalk cancellation system with mismatched sound sources," in Proceedings of the Tenth Western Pacific Acoustics Conference, Beijing, China (2009).