euronoise

Effects of practical loudspeaker characteristics on virtual acoustic imaging systems

Xiaojun Qiu[a] and Michael Vorlaender[b]

[a]Key Laboratory of Modern Acoustics and Institute of Acoustics, Nanjing University, 210093 Nanjing, China

[b]Institute of Technical Acoustics, RWTH Aachen University, Neustr. 50, 52056 Aachen, Germany

xjqiu@nju.edu.cn

Loudspeakers in virtual acoustic imaging systems are usually modeled as monopole sources. This is, however, only an approximation for the low frequency range, since real loudspeaker boxes have a complex directivity in mid to high frequency range. This paper presents an analytical model of crosstalk cancellation systems in free field which takes into account scattering and spatial characteristics of the sound sources. Based on the proposed model, the effects of sound source spatial characteristics on the performance of the crosstalk cancellation system were studied. It was observed that slight sound pressure level difference caused by the sound source directivity might result in large performance loss of the accurate crosstalk cancellation system. The reasons for this performance loss were analyzed for a crosstalk cancellation system configuration with two orientation misaligned directional sound sources. The conclusion is that the performance loss caused by a change on the source span angle seen from the listener is larger than that caused by the sound source directivity.

1 Introduction

Crosstalk cancellation systems (CTC) were proposed nearly 50 years ago and a thorough literature review can be found in recent books and papers [1-4]. It has been found that nearly perfect crosstalk cancellation can be obtained for a matched CTC system, where the playback Head Related Transfer Functions (HRTFs) are the same as the setup HRTFs used to design the CTC filters. However, for a mismatched system, the crosstalk cancellation performance drops as low as 10dB channel separation, even with all HRTFs measured in anechoic rooms [3].

There are many factors that make the playback HRTFs do not match the setup ones. For example, the difference between HRTFs of different individuals with different size and shape of pinnae, head and torso, the misalignment of listener's head position and orientation, the misalignment of playback sound sources' position and orientation, the variation of the spatial and temporal characteristics of the playback sound sources, and the variation of the playback acoustics environment caused by wall reflections.

The performance loss caused by the difference between the setup HRTFs and the playback HRTFs of different individuals has been studied by Akeroyd *et al.* [3]. The study of the CTC robustness to head misalignment by Takashi *et al.* shows that a system with loudspeakers positioned closer to each other is more robust to the misalignment of the listener's head [5]. Rose *et al.* further investigated this for 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 [6]. A study carried out by Bai *et al.* shows that, although the use of a small source span angle of 10° has a large relative sweet spot, larger source span of 60° or even 120° is more desirable in practical applications, as it produces a larger absolute sweet spot [4].

In addition to these robustness studies focused on the listener, there are also some studies on the influences of the playback environments. Ward has derived a closed-form expression to predict the performance of a CTC system in a reverberant environment [7]. The subjective tests carried out by Lentz shows that adding reflecting walls to a listening environment decreases the performance of the CTC system significantly [8]. However, little research has been carried out on the effects of the playback sound sources. The reason might be due to the fact that effects caused by the playback sound sources can usually be made quite small by aligning the playback sound sources more accurately, and by selecting good quality loudspeakers.

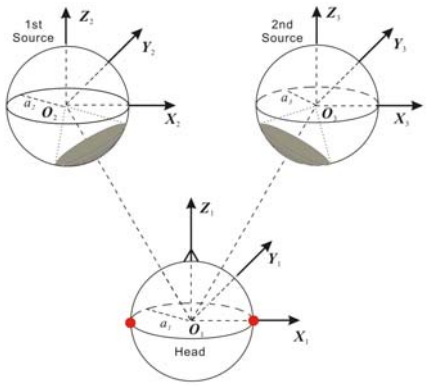
However, in some situations such as in dynamic CTC systems, where the relative angle between loudspeaker and listener is not fixed, the orientation of the playback sources might not be easy to be adequately aligned with the listener, so that the directivity of the sound sources must be taken into account [8]. In some other applications, such as spatial hearing experiments, where high accurate binaural cues are required, it is found that even the scattering and reflections from the loudspeaker cabinets needs to be considered. For example, Akeroyd *et al.* found that the average channel separation of their CTC system can be improved from 20 dB to 23 dB by simply removing the major reflections from the loudspeaker cabinets. These facts motivate the current research to investigate the significance of the effects caused by misalignment of playback sound sources.

Loudspeakers used in CTC systems are usually modelled as monopoles; however, it is an approximation for the low frequency range since real loudspeaker boxes have a complex directivity in mid to high frequency range. An analytical model of CTC systems in free field is developed in this paper, which takes into account scattering and spatial characteristics of sound sources. Then, based on the proposed model, the effects of sound source spatial characteristics on the CTC system are studied.

2 Physical Model

As shown in Fig. 1, there are three spheres (with radius of a_i, i = 1,2,3) in the system, and each 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 translational movement of the original centre from $\mathbf{r}_{o1}$ = (0,0,0) to $\mathbf{r}_{o2}$ = $(d_{o2},\theta_{o2},\phi_{o2})$ and $\mathbf{r}_{o3} = (d_{o3},\theta_{o3},\phi_{o3})$ in coordinate system O_1. The same receiver point is called $\mathbf{r}_{1r} = (r_{1r},\theta_{1r},\phi_{1r})$ in coordinate system O_1 or $\mathbf{r}_{2r} = (r_{2r},\theta_{2r},\phi_{2r})$ in coordinate system O_2 or $\mathbf{r}_{3r} = (r_{3r},\theta_{3r},\phi_{3r})$ in coordinate system O_3.

Sphere 1 is used to approximate 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 polar cap on it. The cap has a span angle of $2\theta_s$ and vibrates radially with a velocity of u_0. This model 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 [9]. Although the error caused by this idealization becomes larger in higher frequency, it can still be used to show the general effects of the loudspeaker directivity and the diffractions on the sound reproduction systems and provide some insights into the basic physics mechanisms.



.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.

The centre of the first polar cap on the surface of Sphere 2 is assumed to be (θ_L , ϕ_L) in its own coordinate system O_2 and, respectively, the second polar cap on the surface of Sphere 3 to be (θ_R, ϕ_R) in its coordinate system O_3. The spherical harmonic coefficients of the velocity distribution of each source in its own coordinate systems can be expressed by [10-14]

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

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

where superscript * denotes the complex conjugate. The time dependent factor $e^{j\omega t}$ is suppressed throughout the analysis, j is the square root of -1 and ω is the angular frequency of interest. The definition of the spherical harmonics is

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

$P_l^m(\cos\theta)$ is the *associated Legendre function* of degree l and order m evaluated at $\cos\theta$. The basic spherical harmonic coefficients

$$U_{l0}^{o} = \frac{u_0}{2}\sqrt{\frac{4\pi}{2l+1}}[P_{l-1}^0(\cos\theta_s) - P_{l+1}^0(\cos\theta_s)] \tag{4}$$

are obtained from the simplest z axis symmetrical case where the centre of the vibrating cap is located at the north pole of the sphere. For $l = 0$, $P_{-1}^0(x) = 1$.

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

$$p_t(\mathbf{r}_{1r}) = p_{s_1}(\mathbf{r}_{1r}) + p_{s_2}(\mathbf{r}_{2r}) + p_{s_3}(\mathbf{r}_{3r}) , \tag{5}$$

$$p_t(\mathbf{r}_{2r}) = p_{s_1}(\mathbf{r}_{1r}) + p_{s_2}(\mathbf{r}_{2r}) + p_{s_3}(\mathbf{r}_{3r}) , \tag{6}$$

$$p_t(\mathbf{r}_{3r}) = p_{s_1}(\mathbf{r}_{1r}) + p_{s_2}(\mathbf{r}_{2r}) + p_{s_3}(\mathbf{r}_{3r}) . \tag{7}$$

It is convenient to express the sound field (radiated and/or scattered) from each sphere in its own coordinate system,

$$p_{s1}(\mathbf{r}_{1r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lm} h_l(kr_{1r}) Y_{lm}(\theta_{1r}, \phi_{1r}) , \tag{8}$$

$$p_{s2}(\mathbf{r}_{2r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} D_{lm} h_l(kr_{2r}) Y_{lm}(\theta_{2r}, \phi_{2r}) , \tag{9}$$

$$p_{s3}(\mathbf{r}_{3r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm} h_l(kr_{3r}) Y_{lm}(\theta_{3r}, \phi_{3r}) , \tag{10}$$

where C_{lm}, D_{lm} and E_{lm} 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, c_0 is the sound speed, $h_m(x) = j_m(x) - j\, n_m(x)$ is the *spherical Hankel function* of order m, $j_m(x)$ is the *spherical Bessel function* of order m and $n_m(x)$ is the *spherical Neumann function* of order m.

The difficulty in the above equations is that there are functions and variables in three different coordinate systems. This can be simplified using the translational addition theorem [12-14], which expresses the sound of the same point in space originally represented by the coordinates in system O_j in terms of the coordinates in coordinate O_i by

$$\begin{aligned} &h_l(kr_{jr}) Y_{lm}(\theta_{jr}, \phi_{jr}) = \\ &\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{oij}) j_p(kr_{ir}) Y_{pq}(\theta_{ir}, \phi_{ir}), \; i, j = 1,2,3. \end{aligned} \tag{11}$$

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

$$\begin{aligned} Q_{pq}^{lm}(\mathbf{r}_{oij}) = &\sum_{n=|l-p|:2}^{l+p} 4\pi(-j)^{p+n-l} h_n(kr_{oij}) Y_{n,m-q}(\theta_{oij}, \phi_{oij}) \\ &\times g(m,l,-q,p,n) \end{aligned} \tag{12}$$

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

$$\begin{aligned} g(m,l,-q,p,n) = &(-1)^m \sqrt{\frac{(2l+1)(2p+1)(2n+1)}{4\pi}} \\ &\times \begin{pmatrix} l & p & n \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & n \\ m & -q & -m+q \end{pmatrix} \end{aligned} \tag{13}$$

is related to the Gaunt coefficients, which is expressed as products of the Wigner 3j. More details on the Wigner 3j symbol and the efficient algorithms for the translational coefficient and Gaunt coefficient calculations can be found in reference [12].

Substitute Eqs. (8-11) into Eqs. (5-7), use the boundary conditions $j\rho_0\omega v_n(\mathbf{r}) = -\partial p_t(\mathbf{r})/\partial r$ at the surface of each rigid sphere in its own coordinate, truncate the number of summations to L (depending on the accuracy required), and equate the coefficients of $Y_{lm}(\theta_{1r},\phi_{1r})$, $Y_{lm}(\theta_{2r},\phi_{2r})$ and $Y_{lm}(\theta_{3r},\phi_{3r})$, the following coupled linear complex equations can be obtained as

$$\begin{aligned} &C_{lm} h_l'(ka_1) + \sum_{p=0}^{L}\sum_{q=-p}^{p} D_{pq} Q_{lm}^{pq}(\mathbf{r}_{o12}) j_l'(ka_1) \\ &\quad + \sum_{p=0}^{L}\sum_{q=-p}^{p} E_{pq} Q_{lm}^{pq}(\mathbf{r}_{o13}) j_l'(ka_1) = 0 \\ &\sum_{p=0}^{L}\sum_{q=-p}^{p} C_{pq} Q_{lm}^{pq}(\mathbf{r}_{o21}) j_l'(ka_2) + D_{lm} h_l'(ka_2) \\ &\quad + \sum_{p=0}^{L}\sum_{q=-p}^{p} E_{pq} Q_{lm}^{pq}(\mathbf{r}_{o23}) j_l'(ka_2) = -j\rho_o c_o U_{lm}^{L} \\ &\sum_{p=0}^{L}\sum_{q=-p}^{p} C_{pq} Q_{lm}^{pq}(\mathbf{r}_{o31}) j_l'(ka_3) + \sum_{p=0}^{L}\sum_{q=-p}^{p} D_{pq} Q_{lm}^{pq}(\mathbf{r}_{o32}) j_l'(ka_3) \\ &\quad + E_{lm} h_l'(ka_3) = -j\rho_o c_o U_{lm}^{R} \end{aligned} \tag{14}$$

where ρ_0 is the density of the medium. The prime denotes differentiation with respect to the argument of the function.

For all l and m (l = 0, 1, 2, …, L, m = –L, –L+1, …, 0, 1, 2, …, L), there are $3(L+1)^2$ unknown spherical harmonic coefficients to be calculated. After they are obtained, the total sound field can be calculated by using any one of Eqs. (5-7). For example,

$$p_t(\mathbf{r}_{1r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} C_{lm} h_l(kr_{1r}) Y_{lm}(\theta_{1r},\phi_{1r}) + \sum_{l=0}^{L}\sum_{m=-l}^{l} D_{lm} h_l(kr_{2r}) Y_{lm}(\theta_{2r},\phi_{2r}) + \sum_{l=0}^{L}\sum_{m=-l}^{l} E_{lm} h_l(kr_{3r}) Y_{lm}(\theta_{3r},\phi_{3r}) \quad (15)$$

Sometimes it is 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. For the radially vibrating polar cap source, $q = 2\pi a_1^2(1-\cos\theta_s)u_0$. 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 Eq. (15) but with the given source spherical harmonic coefficients of Eqs. (1-2) being substituted by,

$$U_{lm}^{L} = \sqrt{\frac{4\pi}{2l+1}}\frac{U_{l0}^{o} Y_{lm}^{*}(\theta_L,\phi_L)}{j\omega\rho_0 a_2^2(1-\cos\theta_s)u_0/2}, \quad (16)$$

$$U_{lm}^{R} = \sqrt{\frac{4\pi}{2l+1}}\frac{U_{l0}^{o} Y_{lm}^{*}(\theta_R,\phi_R)}{j\omega\rho_0 a_3^2(1-\cos\theta_s)u_0/2}. \quad (17)$$

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

3 Simulations and discussion

3.1 Sound source characteristics

Fig. 2 shows the sound source configurations in the simulations where the distance between the point source or source sphere centre and the listener position is 1.0 m, the radius of source spheres is set as $a_1 = a_2 = 0.1$ m, and the radius of listener sphere (head) is $a_3 = 0.09$ m. The sources include a point source, a radially vibrating polar cap on a rigid sphere and a radially vibrating polar cap on a rigid sphere with another rigid sphere aside. When the source is a vibrating cap on the surface of a sphere, the distance is that between the sphere centre and the listener position. The half source span angle of the vibrating cap source θ_s is 30° (about the size of a 4 inch loudspeaker). For the two source cases where another sphere is introduced, the two sources subtend an angle of 60° in Fig. 2(c) or 20° in Fig. 2(d) at the listener position. The location of the vibrating cap centre on the surface of the left sphere (Sphere 2) is (θ_L = 150°, ϕ_L = 0°) in Fig. 2(c) or (θ_L = 170°, ϕ_L = 0°) in Fig. 2(d) in coordinate system O_2, and the location of the vibrating cap on the surface of right sphere (Sphere 3) is (θ_R = 150°, ϕ_R = 180°) in Fig. 2(c) or (θ_R = 170°, ϕ_R = 180°) in Fig. 2(d) in coordinate system O_3.

In a series of preliminary calculations it was found that using a spherical harmonics expansion degree of L = 10 ($(L+1)^2$ = 121 items) can provide sufficient precision with error less than 1 % for the current setup in the current frequency range (up to about 4000 Hz). The higher the ratio of the size of the sphere to the wavelength, the more expansion degrees should be used. In the following simulations, the spherical harmonics expansion degree is set as L = 10. 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 is made for calculation at higher frequency. The frequency response is calculated for source input with the same volume acceleration at 257 points evenly distributed from DC to 4096 Hz.

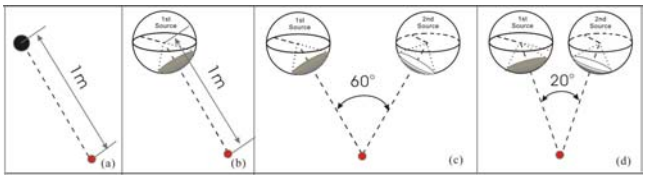


Fig. 2 Configurations of the sound sources

Fig. 3 shows the magnitude of the calculated Frequency Response Functions (FRF) in dB (ref. 1.0) for the configurations in Fig. 2 and a baffled piston with a radius of 0.1 m. Note that for the configurations in Figs. 2(c-d), the second source on the right sphere is not active and the sphere is assumed rigid, so the frequency response is just that of one source. It can be seen that the sources of a vibrating cap on a rigid sphere, as shown in Figs. 2(b-d), do not have a flat frequency response as does the ideal point source of Fig. 2(a). The vibrating cap on a rigid sphere radiates like a point source in the low frequency range (the slight difference around DC is due to the distance difference from the sources to the receiver, which is less than 1.0 m for the vibrating cap case), and radiates two times higher pressure in high frequency range like the baffled piston. This is caused by the scattering of rigid sphere where the cap sound source locates.

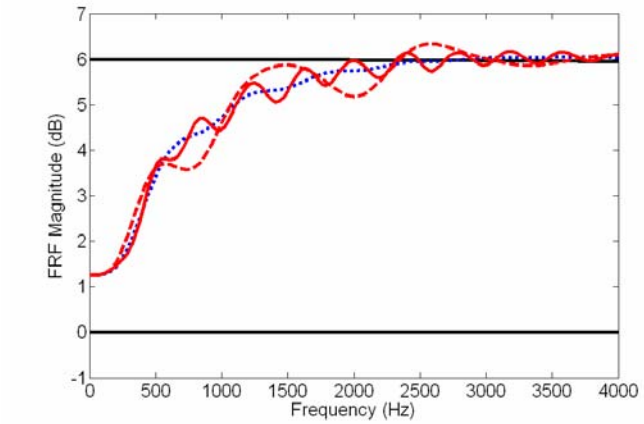


Fig. 3 The magnitude of the calculated FRFs for each configuration of Fig. 2. The lower black solid line is for 2(a), the upper black solid line is that of the baffled piston, the blue dotted line is for Fig. 2(b), the red solid line is for Fig. 2(c), and the red dashed line is for Fig. 2(d).

It can also be observed from the curves for Figs. 2(c-d) that the scattering from a surrounding rigid sphere adds some fluctuations on the original frequency response. The magnitude and frequency of the fluctuations depends on the

distance between the two spheres. The nearer the spheres, the larger the fluctuation magnitude is. Another interesting phenomenon is that the magnitude of the fluctuation becomes smaller at higher frequency and the reason for this is that the cap source behaves more directional at higher frequency so that less sound radiates to and less reflection and scattering comes from the sphere aside.

Fig. 4 shows the calculated responses at 2000 Hz when the receiving point swings around the source from –30° (left) to 30° (right) for configurations of Fig. 2 (current line along the source and the receiving point is 0° line and the distance remains unchanged). Although these curves are not the directivities of the sound sources in the far field, they are given here to show the difference of sound field spatial distributions between these configurations. It can be seen that the vibrating cap on a rigid sphere shows certain directivity and the magnitude at 30° can be 2 dB lower than that on the centre axis. The scattering from aside rigid sphere makes the spatial distribution rather complicated.

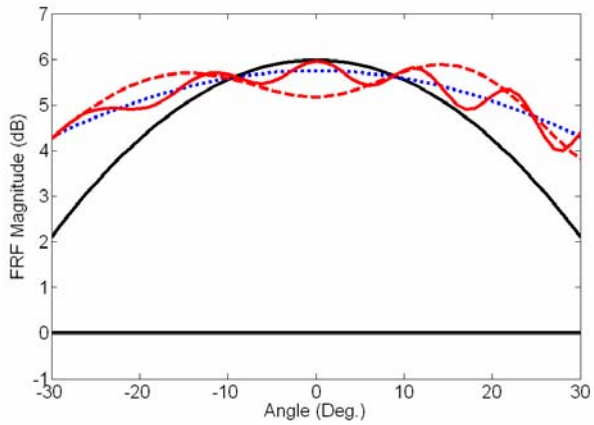


Fig. 4 The magnitude of calculated FRFs at 2000 Hz when the receiving point swings from –30° (left) to 30° (right) for configurations of Fig. 2. The lower black solid line is for Fig. 2(a), the upper black solid line is that of the baffled piston, the blue dotted line is for Fig. 2(b), the red solid line is for Fig. 2(c), and the red dashed line is for Fig. 2(d).

3.2 CTC with non-ideal sound sources

Fig. 5 shows the block diagram of a CTC system where **C** is a 2×2 plant transfer function matrix, **V** is the source input signal vector, **U** is the recorded signal vector, and **H** is the 2×2 CTC matrix, and can be obtained by

$$\mathbf{H} = [\mathbf{C}^{\mathrm{H}}\mathbf{C} + \beta\mathbf{I}]^{-1}\mathbf{C}^{\mathrm{H}}e^{-j\omega\Delta}, \qquad (18)$$

where β is a regularization parameter, being used to constraint the energy of the source input signals. The performance of a cross-talk cancellation system is described by the performance matrix below,

$$\mathbf{P} = \mathbf{CH}, \qquad (19)$$

where **P** is a unit matrix under ideal crosstalk cancellation, with its element $P_{11} = P_{22} = 1$ and $P_{12} = P_{21} = 0$. The above matrixes **C**, **H** and **P** are functions of frequency. To show the performance variation with frequency, a single value CTC performance index is often used. This paper uses the channel separation, as does references [4, 5], given by

$$CHSP(\omega) = 20\log_{10}(|P_{11}(\omega)/P_{12}(\omega)|) \qquad (20)$$

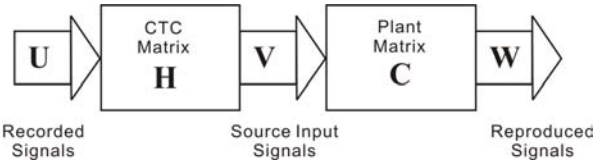


Fig. 5 Block diagram of a crosstalk cancellation system

Fig. 6 shows three analytical models used in the CTC system study, which has the same geometry configurations as that mentioned in Fig. 2. The commonly used models are the free field model, shown in Fig. 6(a), and the spherical head model, shown in Fig. 6(b). Sometimes, the HRTFs model are used for more comprehensive study and practical CTC system design where the transfer functions from source to the rigid sphere of Fig. 6(b) are substituted by measured HRTFs from dummy heads or individual persons. These two models assume the playback sound sources as ideal point sources. Fig. 6(c) shows the model applied in this paper that can be used to study the non-ideal sound sources and their interaction.

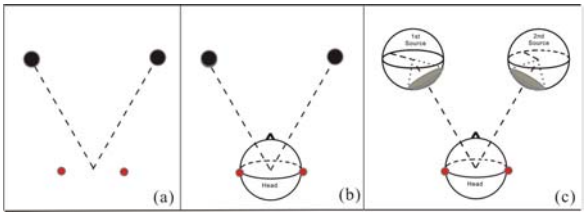


Fig. 6 The analytical models for the crosstalk cancellation system studies

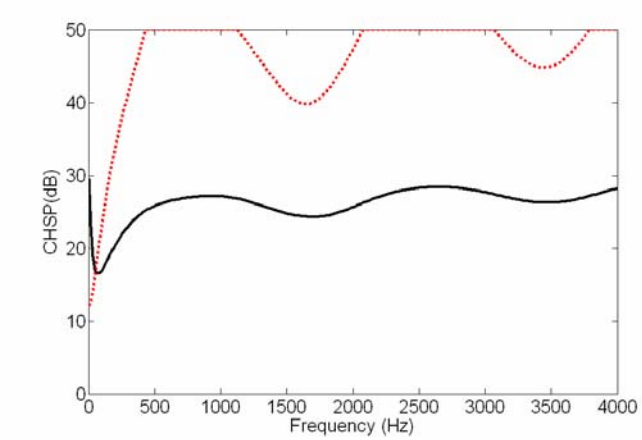


Fig. 7 Channel separation of the CTC system using different playback plant transfer functions with 60° source span. The setup plant transfer functions are obtained with the point sound sources of Fig. 6(b). The playback sound sources are the same as the setup ones (red dotted line) or the ones that radiate smaller in the crosstalk path direction with the transfer function magnitude 0.9 times of that in the direct path (black solid line).

Model Fig. 6(b) is used to study the effects of sound source directivity on CTC performance by assigning different gains at different directions on its playback plant transfer functions. The use of this simple model can emphasize the effects of the sound source directivity. The directivity of the playback plant transfer function are simulated by directly multiplying a small value of 0.9 on the crosstalk path transfer functions, which corresponds to about 0.9 dB excess sound pressure level difference between two ears.

This simple approximation might be too rough to model the frequency dependent directivity of real loudspeakers; however, it clearly shows that such a small magnitude difference between setup and playback plant transfer functions might reduce channel separation from more than 40 dB to about 25 dB as shown in Fig. 7. In the simulations, the regularization parameter β is adjusted so that the channel separation under ideal situation (the setup plant transfer functions are also used as the playback ones) is greater than 40 dB above 500 Hz, and channel separation greater than 50 dB is clipped at 50 dB throughout the paper.

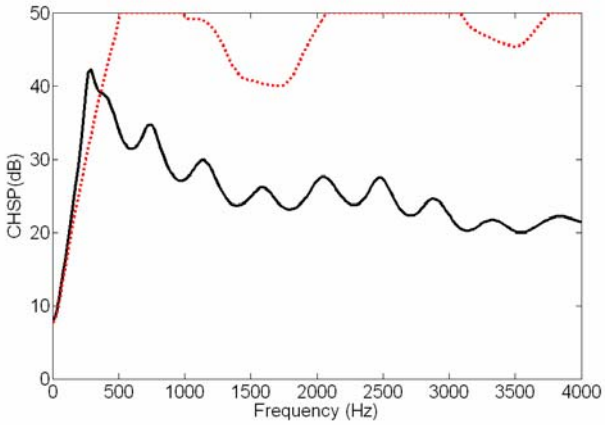


Fig. 8 Channel separation of the CTC system using different playback plant transfer functions with 60° source span. The setup plant transfer functions are obtained with the vibrating cap sound sources of Fig. 6(c). The playback sound sources are the same as the setup ones (red dotted line) or obtained by rotating the cap 10° away from its own original source central axis (black solid line).

Another example is to change the orientation of the playback sound sources so that it is misaligned by 10° away from its own original source central axis. After rotation, the location of the vibrating cap on the surface of the left sphere (Sphere 2) is (θ_L = 160°, ϕ_L = 0°) in Fig. 2(c) in coordinate system O_2, and that of right sphere (Sphere 3) is (θ_R = 160°, ϕ_R = 180°) in Fig. 2(c) in coordinate system O_3. It can be seen in Fig. 8 that the channel separation drops significantly from more than 40 dB to about 20 dB. Detailed investigations (not shown here) found that there are four contributors for the performance loss: the slight change of the distance from sources to the listener, the directivity of the playback sound sources, the scattering from the aside sphere, and the misalignment of the subtended angle of the playback sound sources seen from the listener, where the last factor plays a key role.

4 Conclusion

The commonly used analytical models in virtual acoustic imaging studies are the free field model and the spherical head model where the sound sources are assumed as point sources. This paper develops an analytical model based on spherical harmonics decomposition and the translational addition theorem where the scattering and spatial characteristics of the sound sources are considered. The model is general and can be extended to multiple listeners and loudspeakers cases. Based on the proposed model, the effects of sound source spatial characteristics on crosstalk cancellation systems are studied. It is found that slight sound pressure level difference between two ears caused by the sound source directivity might cause significant performance loss to an accurate crosstalk cancellation system. For a CTC configuration with two orientation misaligned directional sound sources, several factors for its performance loss are analyzed, and it is found that the main contributor for the performance loss at most frequencies comes from the misalignment of the subtended angle of the playback sound sources seen from the listener in this case.

Acknowledgments

The first author is a research fellow of the Alexander von Humboldt Foundation while carrying out this research which is partially supported by Project 10674068 of NSFC and NCET, and the third author is a Scholarship holder from CNPq, Brazil.

References

[1] M. Vorländer, *Auralization*, Springer. (2007)
[2] P. A. Nelson and J. F. W. Rose, "The time domain response of some systems for sound reproduction", *J. Sound Vib.*, 296(3), 461-493 (2006)
[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(4), 1976-1989 (2006)
[5] T. Takeuchi and P. A. Nelson, and H. Hamada, "Robustness to head misalignment of virtual sound imaging systems", *J. Acoust. Soc. Am.* 109 (3), 958-971 (2001)
[6] J. Rose, P. A. Nelson, B. Rafaely, and T. Takeuchi, "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations", *J. Acoust. Soc. Am.* 112, 1992-2002 (2002)
[7] D. B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment", *J. Acoust. Soc. Am.* 110 (2), 1195-1198 (2001)
[8] T. Lentz, *Binaural technology for virtual reality*. Doctoral thesis, Institute of Technical Acoustics, RWTH Aachen University, Germany. (2007)
[9] J. Garcia-Bonito and S. J. Elliott, "Local active control of diffracted diffuse sound fields", *J. Acoust. Soc. Am.*, 98(2), 1017–1024 (1995)
[10] E. G. Williams, *Fourier Acoustics*, Elsevier. (1999)
[11] M. Pollow, *Variable directivity of dodecahedron loudspeaker*s, Diploma theses, Institute of Technical Acoustics, RWTH Aachen Univ., Germany. (2007)
[12] W. C. Chew, "Recurrence Relations for Three-Dimensional Scalar Addition Theorem", *J. Electromag. Waves and Appl.* 6(2), 133-142 (1992)
[13] G. C. Gaunaurd, H. Huang, and H. C. Strifors, "Acoustic scattering by a pair of spheres", *J. Acoust. Soc. Am.*, 98(1), 495-507 (1995)
[14] H. Huang and G. C. Gaunaurd, "Acoustic scattering of a plane wave by two spherical elastic shells", *J. Acoust. Soc. Am.*, 98(4), 2149-2156 (1995)