Cross-talk Cancellation Algorithm for 3D Sound Reproduction

Hyoun-Suk Kim[a], Poong-Min Kim, and Hyun Bin Kim

If the right and left signals of a binaural sound recording are reproduced through loudspeakers instead of a headphone, they are inevitably mixed during their transmission to the ears of the listener. This degrades the desired realism in the sound reproduction system, which is commonly called 'cross-talk.' A 'cross-talk canceler' that filters binaural signals before they are sent to the sound sources is needed to prevent cross-talk. A cross-talk canceler equalizes the resulting sound around the listener's ears as if the original binaural signal sound is reproduced next to the ears of listener. A cross-talk canceler is also a solution to the problem—how binaural sound is distributed to more than 2 channels that drive sound sources. This paper presents an effective way of building a cross-talk canceler in which geometric information, including locations of the listener and multiple loudspeakers, is divided into angular information and distance information. The presented method makes a database in an off-line way using an adaptive filtering technique and Head Related Transfer Functions. Though the database is mainly concerned about the situation where loudspeakers are located on a standard radius from the listener, it can be used for general radius cases after a distance compensation process, which requires a small amount of computation. Issues related to inverting a system to build a cross-talk canceler are discussed and numerical results explaining the preferred configuration of a sound reproduction system for stereo loudspeakers are presented.

I. INTRODUCTION

The overwhelming part of current research in virtual sound source imaging relies heavily on binaural technology. Binaural technology is deeply related to Head Related Transfer Functions (HRTFs) [1]–[5], and is based on the principle that if a sound reproduction system can generate the same sound pressures at the listener's eardrums as is produced there by a real sound source, then the listener can not tell the difference between the virtual image and the real sound source. In order to understand these binaural signals, or "target" signals, it is necessary to understand how the listener's torso, head, and pinnae (outer ears) modify incoming sound waves as a function of the position of the sound source. This information which is called HRTFs can be obtained by making measurements on "dummy-heads" or human subjects. One can create sonic illusions by using complex filters using the HRTFs. Sonic illusion is a virtual sound source that does not exist in the real world. The listener can artificially conceive it as floating in mid air or rising up in front of the listener if sound signals driving the real sound sources are controlled in a complicated way. Virtual source imaging systems work on the principle that they "get the sound right" at the ears of the listener. In practice, loudspeakers are used to reproduce a set of "desired" signals in the region around the listener's ears. The desired signals contain the sound "emitted" from the virtual source as well as their spatial information. The inputs to the loudspeakers must be recombined from the desired signals according to the relative geometric configuration of speakers and listener.

One of the popular sound reproduction systems is the transaural audio system that delivers binaural signals to the ears of a listener using stereo loudspeakers. The basic idea is to filter the binaural signal so that the subsequent stereo presentation

Manuscript received July 7, 1999; revised April 11, 2000.
[a] Electronic mail: kimmhs@etri.re.kr

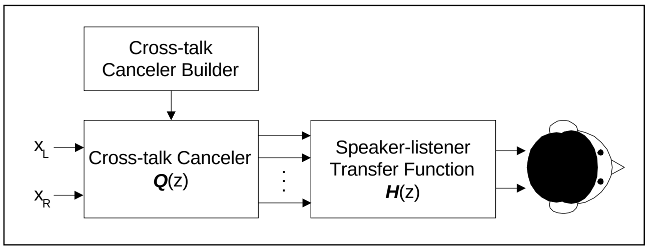

Fig. 1. Binaural sound reproduction including cross-talk canceler.

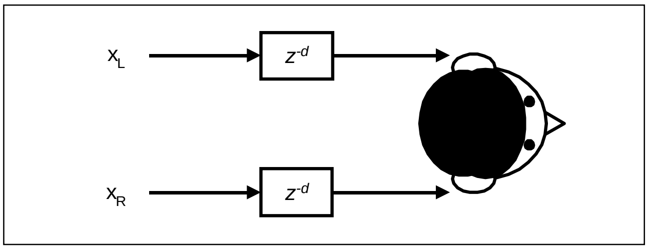

Fig. 2. Equivalent system between binaural sound and listener with cross-talk canceler.

produces the desired binaural signal at the ears of the listener. Binaural-type two-channel sound that is created using sound localization or a virtual sound field control technique fits well for headphone reproduction systems. A stereo headphone transmits left and right channels direct to the listener's left and right ears, respectively, with negligible cross-mixing in general. If two-channel binaural sound is reproduced through stereo loudspeakers instead, sound received by the listener can be severely changed from the original binaural sound due to the acoustic system between the sound sources (loudspeakers) and the listener. The emitted sound from the left loudspeaker goes to the right ear as well as to the left ear of the listener, and vice versa.. Thus, a simple stereo loudspeaker reproduction system cannot deliver the realism of original binaural sound without appropriate preprocessing. The paths from the right loudspeaker to the left ear and vice versa are called cross-talk paths. The effect of cross-talk can be avoided if both channels in binaural sound are properly recombined or filtered before they are sent to the stereo loudspeakers [6]–[8].

When more than two loudspeakers are used for reproducing a binaural sound, one should be concerned with the problem of how the binaural sound is mapped into more than two sound signals. This can also be solved following the same process of recombining or filtering binaural signals as the cross-talk canceler does. If there are *S* sound sources, both problems (constructing the cross-talk canceler and recombining/mapping binaural signals into multiple signals) can be considered as the same problem of designing a system whose input is a 2-channel binaural sound and whose output is a *S*-channel sound that drives the loudspeakers.

This paper assumes the HRTFs are generally measured in an anechoic chamber and all acoustic components around the listener, including acoustic properties of the listening room and the reproduction system, are close to those of where HRTFs are measured. These assumptions simplify an ordinary sound reproduction system, denoted by $\boldsymbol{H}(z)$, with HRTFs only. Also it is assumed that every amplifier channel is equal and every response of the loudspeakers is almost the same.

Our approach to build a cross-talk canceler is as follows. First, build a cross-talk canceler database for full or some selected combination of discretized angles of loudspeakers assuming they are on a *standard radius* from the listener. The standard radius is the distance from the listener to the loudspeaker when HRTFs are measured. General cases where the loudspeakers are not on the standard radius can be obtained from the relevant cross-talk filter stored in the database, using a distance compensation process. The distance compensation includes two kinds of compensation:gain compensation and delay compensation, which are numerically quite simple. The following sections will describe the detailed procedure of the approach starting with an *S* loudspeaker system and apply it to the simplest system (stereo loudspeaker system) incorporating some novel and efficient algorithms. The problem arising while solving the inverse of $\boldsymbol{H}(z)$ to build the cross-talk canceler is investigated and an 'invertability measure' is suggested to measure the accuracy or ease of inverting $\boldsymbol{H}(z)$.

II. SOUND REPRODUCTION SYSTEM WITH CROSS-TALK CANCELER

Consider the binaural sound reproduction system with *S* loudspeakers. The cross-talk canceler is usually placed before the ordinary reproduction system $\boldsymbol{H}(z)$ which is an *S*-input 2-output system as depicted in Fig. 1. Let's denote the cross-talk canceler by $\boldsymbol{Q}(z)$. Its input is a binaural signal and its output is *S* signals that are designated to drive the corresponding loudspeakers. In this paper, the aim of the cross-talk canceler is to convert the combined system of itself and $\boldsymbol{H}(z)$ to a simple system that includes an endurable amount of delay as follows:

$$\boldsymbol{H}(z)\boldsymbol{Q}(z) = z^{-d}\boldsymbol{I}. \quad (1)$$

If $\boldsymbol{Q}(z)$ is well designed so that (1) is satisfied, the whole reproduction system is converted to a simple delay system as illustrated in Fig. 2, and the realism of the binaural sound will be transfered to the listener with less deterioration. The delaying term, z^{-d}, is necessary because it is impossible to get a real-time applicable $\boldsymbol{Q}(z)$ when $\boldsymbol{H}(z)$ is a nonminimum phase system (true for common audio reproduction systems).

The ordinary reproduction system, $\boldsymbol{H}(z)$, is the product of two transfer functions $\boldsymbol{T}(z)$ and $\boldsymbol{M}(z)$, where $\boldsymbol{T}(z)$ is the simplified loudspeaker-listener transfer function when speakers are on a standard radius, which is expressed in terms of the HRTFs as in (3), and $\boldsymbol{M}(z)$, is a diagonal matrix compensating the distance variation from the standard radius. Such a simple approximation of $\boldsymbol{H}(z)$ is true for an anechoic chamber, but it is not true in general. However, this approximation is reasonable if absorption coefficients of the room walls are high enough.

$$\boldsymbol{H}(z) = \begin{bmatrix} H_{L1}(z) & H_{L2}(z) & \cdots & H_{LS}(z) \\ H_{R1}(z) & H_{R2}(z) & \cdots & H_{RS}(z) \end{bmatrix}$$

$$= \begin{bmatrix} H_L(\theta_1, z) & H_L(\theta_2, z) & \cdots & H_R(\theta_S, z) \\ H_R(\theta_1, z) & H_R(\theta_2, z) & \cdots & H_R(\theta_S, z) \end{bmatrix}$$

$$\times \begin{bmatrix} m(r_1, r_0, z) & & & 0 \\ & m(r_2, r_0, z) & & \\ & & \ddots & \\ 0 & & & m(r_S, r_0, z) \end{bmatrix}$$

$$= \boldsymbol{T}(z)\boldsymbol{M}(z), \quad (2)$$

$H_L(\theta_i, z)$: HRTF from *i*-th loudspeaker at angle θ_i to left ear, (3-a)

$H_R(\theta_i, z)$: HRTF from i-th loudspeaker at angle θ_i to right ear, (3-b)

$$m(r_i, r_0, z) = \frac{r_0}{r_i} z^{-(r_i - r_0) f / c}, \quad (4)$$

where, the subscript *R* and *L* stand for right and left, respectively, and $m(r_i, r_0, z)$ stands for the changes of the transfer function caused by the distance variation r_i from the standard radius r_0. The distance compensator, $m(r_i, r_0, z)$, compensates the gain and delay compared to the standard-radius case. The quantity *c* is the speed of sound and *f* is the sampling frequency of the sound to be reproduced.

Using (2), and the design restriction of $\boldsymbol{Q}(z)$ in (1), it is easy to show that the cross-talk canceler $\boldsymbol{Q}(z)$ is the product of $\boldsymbol{M}(z)^{-1}$ and $\boldsymbol{C}(z)$, where $\boldsymbol{C}(z)$ is at this time the delayed-inverse of $\boldsymbol{T}(z)$.

$$\boldsymbol{Q}(z) = \boldsymbol{M}(z)^{-1}\boldsymbol{C}(z), \quad (5)$$

$$\boldsymbol{T}(z)\boldsymbol{C}(z) = z^{-d}\boldsymbol{I}, \quad (6\text{-a})$$

$$\boldsymbol{M}(z)^{-1} = \begin{bmatrix} m(r_0, r_1, z) & & & 0 \\ & m(r_0, r_2, z) & & \\ & & \ddots & \\ 0 & & & m(r_0, r_S, z) \end{bmatrix}. \quad (6\text{-b})$$

The *i*-th diagonal term of $\boldsymbol{M}(z)^{-1}$ is simply $m(r_0, r_i, z)$, the inverse of $m(r_i, r_0, z)$.

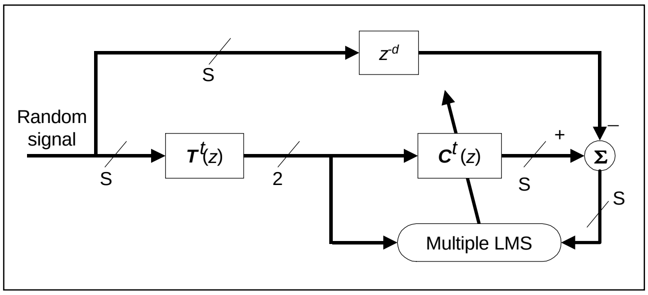

Fig. 3. Adaptation of *C*(z).

The $\boldsymbol{C}(z)$, satisfying $\boldsymbol{T}(z)\boldsymbol{C}(z) = z^{-d}\boldsymbol{I}$, can be obtained using the Multiple Filtered-X LMS algorithm[7]–[9]. This algorithm adapts $\boldsymbol{C}(z)$ to minimize the difference of $\boldsymbol{T}(z)\boldsymbol{C}(z)$ and $z^{-d}\boldsymbol{I}$ in the Least Mean Squares sense with an adaptive FIR model by modeling the 2*S* internal SISO transfer function in $\boldsymbol{C}(z)$. This algorithm suffers from slow convergence speed and requires a large amount of calculation. If $\boldsymbol{T}(z)$ is not well conditioned or its delay becomes larger, the required time to make *C*(*z*) converge increases. This difficulty can be alleviated by converting the restriction condition of (6-a) to $\boldsymbol{C}'(z)\boldsymbol{T}'(z) = z^{-d}\boldsymbol{I}$; applying the transpose operation as done by Kim [10]. Then $\boldsymbol{C}(z)$ can be adapted with the Multiple LMS that is much simpler than Multiple Filtered-X LMS algorithm. Figure 3 is a schematic diagram showing how $\boldsymbol{C}(z)$ is adapted with the Multiple LMS algorithm.

When $\boldsymbol{T}(z)$ changes, $\boldsymbol{C}(z)$ must be readapted; changes of relative position between the loudspeakers and the listener require recalculation of $\boldsymbol{C}(z)$. However, the speaker position along the radial direction can be accommodated by the matrix $\boldsymbol{M}(z)$. This is why it is more effective to build a database of $\boldsymbol{C}(z)$ (the standard radius cases) than $\boldsymbol{Q}(z)$ (whole cases). The database of the standard radius cases can be extended to any loudspeaker locations, provided that the distance between the listener and speaker is not too close to allow the approximation of $\boldsymbol{H}(z) = \boldsymbol{T}(z)\boldsymbol{M}(z)$ to hold. The compensation of distance variation from the standard radius is accomplished simply by multiplying $\boldsymbol{M}(z)^{-1}$. The inversion problem in getting $\boldsymbol{C}(z)$ may introduce some controversial issues, which will be discussed in section IV.

Figure 4(a) shows an example configuration using a three-loudspeaker reproduction system. The projection of loudspeakers on the standard radius is shown in Fig. 4(b). If one has a cross-talk canceler for the configuration of Fig. 4(b), any cross-talk canceler related to the reproduction system having the same projected configuration can be derived from it; Fig. 4(a) is just one of such infinite configurations derivable. If the loudspeaker is very close to the listener, however, correction of the HRTFs is needed.

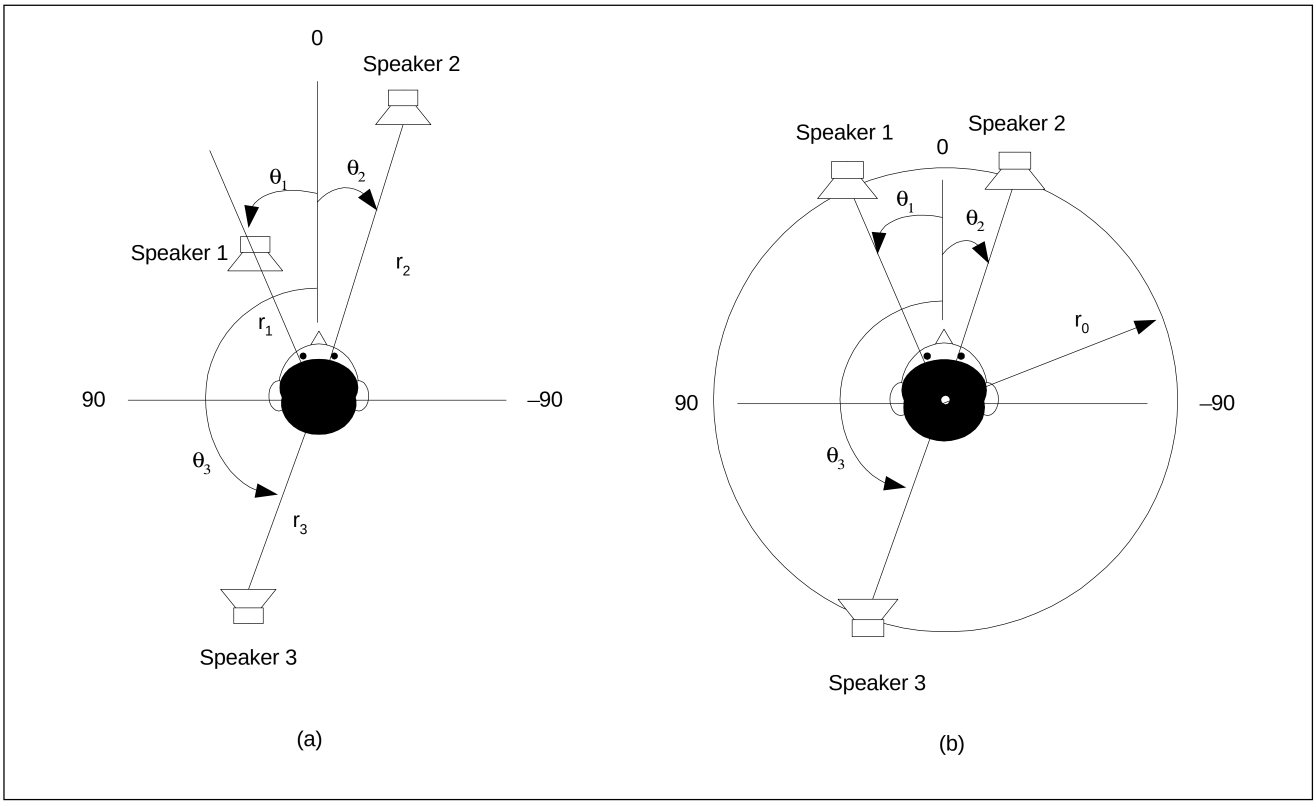

Fig. 4. (a) A reproduction system with $S=3$ and (b) its projection on the standard radius.

III. STEREO LOUDSPEAKER SYSTEM

It is most common to use stereo loudspeakers in a sound reproduction system. Figure 5 shows a transaural audio system using such stereo loudspeakers with a cross-talk canceler. $\boldsymbol{C}(z)$, the delayed-inverse of $\boldsymbol{T}(z)$, can be recursively computed, for example, using the Multiple LMS (or Multiple Filtered-X LMS) algorithm. For the two loudspeaker reproduction system, however, there exists a simpler way of adapting $\boldsymbol{C}(z)$ as presented below.

From (2) and (5), a 2-input, 2-output delayed-inverse $\boldsymbol{C}(z)$ is rearranged as

$$\boldsymbol{C}(z)=\begin{bmatrix} C_{11}(z) & C_{12}(z) \\ C_{21}(z) & C_{22}(z) \end{bmatrix} = \frac{z^{-d}\begin{bmatrix} H_L(\theta_1,z) & -H_R(\theta_1,z) \\ -H_L(\theta_2,z) & H_R(\theta_2,z) \end{bmatrix}}{H_L(\theta_1,z)H_R(\theta_2,z)-H_L(\theta_2,z)H_R(\theta_1,z)}. \quad (7)$$

Denoting the delayed-inverse of the determinant of $\boldsymbol{T}(z)$ by $Q_0(z)$, $\boldsymbol{C}(z)$ becomes

$$\boldsymbol{C}(z)=Q_0(z)\begin{bmatrix} H_L(\theta_1,z) & -H_R(\theta_1,z) \\ -H_L(\theta_2,z) & H_R(\theta_2,z) \end{bmatrix}, \quad (8)$$

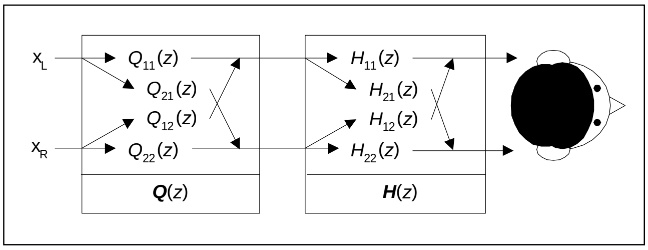

Fig. 5. Binaural sound reproduction using 2 loudspeakers.

$$Q_0(z)=\frac{z^{-d}}{H_L(\theta_1,z)H_R(\theta_2,z)-H_L(\theta_2,z)H_R(\theta_1,z)} = \frac{z^{-d}}{|\boldsymbol{T}(z)|}. \quad (9)$$

Since the $Q_0(z)$ is a SISO system, the LMS algorithm is sufficient for its update. Following this approach, one can reduce the computational effort by about 75%. Figure 6 shows the LMS implementation for adapting $Q_0(z)$. This minimizes the difference between $Q_0(z)\,|\boldsymbol{T}(z)|$ and z^{-d} in the Least Mean Square sense.

If $Q_0(z)$ s are obtained for a popular set of (θ_1,θ_2), we can build a database of $\boldsymbol{C}(z)$ for the set using (8). Figure 7 summarizes the detailed steps to build a cross-talk canceler for

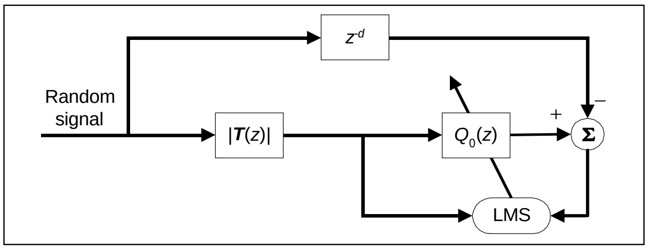

Fig. 6. Adaptation of Q_0(z).

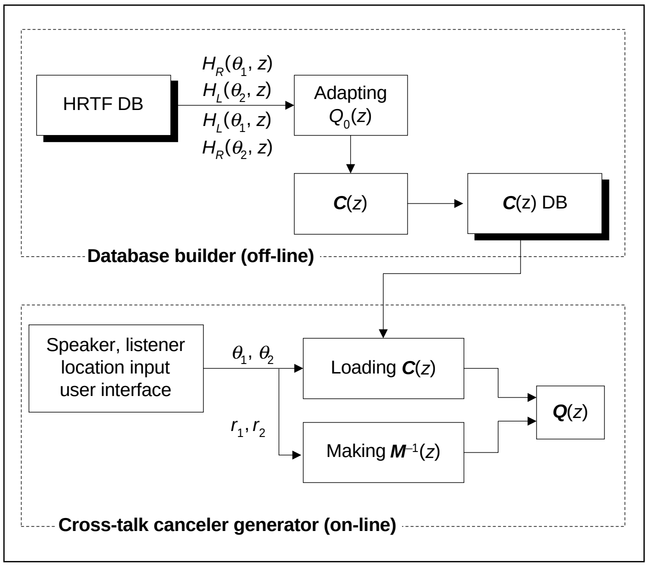

Fig. 7. Constructing a cross-talk canceler ($S=2$).

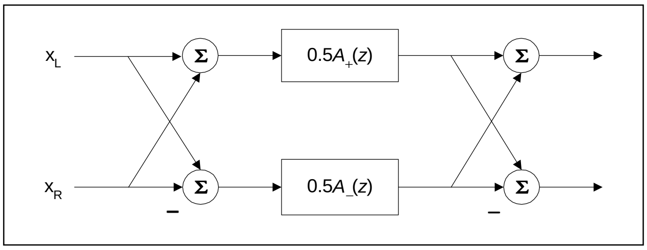

Fig. 8. Shuffler implementation of transaural filter **C**(z) for a symmetric arrangement.

the stereo reproduction system. This comprises of two blocks. The first block is concerned with building the database of $\boldsymbol{C}(z)$ using HRTFs and the adaptive filter technique; the second block generates a cross-talk canceler in real-time using the database of $\boldsymbol{C}(z)$. It is straightforward to calculate the cross-talk canceler $\boldsymbol{Q}(z)$ for the general radius case after $\boldsymbol{C}(z)$ is achieved (see (5), (6-b)).

• *Symmetric Angle Listening Situation*

It is quite common to have loudspeakers of the stereo reproduction system arranged symmetrically from the listener's viewpoint. When listening situation is symmetric in angle, $\boldsymbol{C}(z)$ in (8) can be specified in terms of the ipsilateral $(C_i(z) = C_{11}(z) = C_{22}(z))$ and contralateral $(C_c(z) = C_{12}(z) = C_{21}(z))$ responses. In this case, one can use the 'shuffler' implementation [11] of the transaural filter which involves forming the sum and difference of two input channels to $\boldsymbol{C}(z)$, filtering these signals, and then undoing the sum and difference operations. The sum and difference operations are accomplished by the unitary matrix **D**. The shuffler matrix diagonalizes the matrix $\boldsymbol{C}(z)$ via a similarity transformation as follows:

$$\boldsymbol{C}(z) = \begin{bmatrix} C_i(z) & C_c(z) \\ C_c(z) & C_i(z) \end{bmatrix} = \boldsymbol{D}^{-1} \begin{bmatrix} A_+(z) & 0 \\ 0 & A_-(z) \end{bmatrix} \boldsymbol{D}, \tag{10}$$

$$\boldsymbol{D} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tag{11}$$

$$A_+(z) = C_i(z) + C_c(z), \tag{12-a}$$

$$A_-(z) = C_i(z) - C_c(z), \tag{12-b}$$

where $A_+(z)$ and $A_-(z)$ are the sum and the difference of $C_i(z)$ and $C_c(z)$. Note that **D** is its own inverse, and the normalizing gain $1/\sqrt{2}$ can be commuted to a single gain 1/2. Shuffler implementation helps reconstruction of $\boldsymbol{C}(z)$ to a simpler structure as shown in Fig. 8. The shuffler implementation reduces the convolution operation required in filtering through $\boldsymbol{C}(z)$ by half, although 4 additions and 2 multiplications (alternatively, a shift operation can be used) are needed for each sample output.

IV. DISCUSSION: INVERSE PROBLEM

1. Performance of cross-talk cancellation

There are four major causes of imperfect cancellation of the cross-talk. The first one is singularity in $\boldsymbol{T}(z)$ due to the common zeros in $\boldsymbol{T}(z)$. The common zeros in the every SISO channel in $\boldsymbol{T}(z)$ make it singular. The inevitable low gain of a loudspeaker used as a control source at very low frequency prohibits the control output from going through $\boldsymbol{T}(z)$ and generates common zeros. Therefore, perfect cross-talk cancellation is not achievable in these very low frequency regions where $\boldsymbol{T}(z)$ has common zeros. The second cause of imperfect cancellation is the singularity in $\boldsymbol{T}(z)$ due to bad location of the sound sources. If the distance between sound sources is too short compared with the wavelength of the sound to be reproduced, $\boldsymbol{T}(z)$ becomes almost singular. For a certain geometry of sound sources and listener, rows or columns of $\boldsymbol{T}(z)$ may be linearly dependent and $\boldsymbol{T}(z)$ can be singular in some frequency ranges. In such cases, $\boldsymbol{T}(z)$ can be made non-singular by readjusting the position of sound sources and listener. At a very low frequency range near zero, however, wavelength becomes very long compared with the distance between sound sources and listener, and it becomes singular without exception. The third cause is erroneous modeling of $\boldsymbol{T}(z)$, and the fourth cause is erroneous adaptation of $\boldsymbol{C}(z)$, which can be caused by assigning the length of the $\boldsymbol{C}(z)$ filter too short or by the improper value of d.

Accurate cancellation of the cross-talk is possible, if $\boldsymbol{T}(z)$ is ideally conditioned, its model is accurate for the frequency range of interest, and the order of $\boldsymbol{C}(z)$ is assigned sufficiently high with appropriate d. It is worth noting that if $\boldsymbol{T}(z)$ has common zeros for a certain frequency, it is obvious that there will be no sound power transmission for that frequency even with the help of the cross-talk canceler. Hence, achieving the inverse of $\boldsymbol{T}(z)$ for the specified frequency is not necessary because it cannot alter the sound at the ears of the listener. To save the effort of loudspeaker of the reproduction system, it is recommended that the excitation signal used in the adaptation process should be a random signal excluding frequency regions where $\boldsymbol{T}(z)$ has common zeros.

Note again that $\boldsymbol{C}(z)$ is closely related to the inverse of $\boldsymbol{T}(z)$ (the condition of the equation can be alternatively expressed as $\boldsymbol{C}(z)=\boldsymbol{T}^{-1}(z)z^{-d}$). If $\boldsymbol{T}(z)$ is not *diagonally dominant*, $\boldsymbol{C}(z)$ becomes very sensitive to the variation of $\boldsymbol{T}(z)$ and small changes in $\boldsymbol{C}(z)$ may deviate $\boldsymbol{T}(z)\boldsymbol{C}(z)$ very far from $z^{-d}\boldsymbol{I}$. Such a mathematical property strongly suggests that $\boldsymbol{T}(z)$ is constructed as diagonally dominant as possible. From a physical point of view, $\boldsymbol{T}(z)$ becomes diagonally dominant if each sound source is close to the listener's pinnae. An extreme configuration for such a case is the headphone system, which is best for minimizing the problem that may occur when the cross-talk canceler is applied.

2. Invertability Index

The adaptive scheme described in Fig. 3 converges very slowly at the singular frequency of $\boldsymbol{T}(z)$, because the output of $\boldsymbol{T}'(z)$ will be very small and it takes a long time for such a weak signal to magnify the response of $\boldsymbol{C}'(z)$ in the updating process. Even if a good and accurate inverse model of $\boldsymbol{C}(z)$ were obtained, there would be a significant problem remaining. At the frequency where $\boldsymbol{T}(z)$ becomes singular, the response of $\boldsymbol{C}(z)$ should go to infinity to make the gain of their product unity, and consequently the output of the cross-talk canceler may go over the maximum input range of the amplifier circuit. One finally encounters a dilemma that keeping both accurate inverse modeling and a moderate level of signal to the amplifier at the same time is impossible. Since such a problem originates from the singularity of $\boldsymbol{H}(f)$, the frequency response of $\boldsymbol{H}(z)$, it is necessary to consider an invertability measure (*IM*) of $\boldsymbol{H}(f)$ at the frequency *f*. The simplest representative value of $\boldsymbol{H}(f)$ is its determinant. However, it is not appropriate for IM, because even scaling of $\boldsymbol{H}(f)$ results in a change of the IM. An appropriate measure is the ratio of the minimum to maximum singular value of $\boldsymbol{H}(z)$. This is the ratio of the smallest amplification factor to the largest amplification factor, which is the inverse of the condition number of matrix $\boldsymbol{H}(f)$.

$$IM(f)=\frac{\sigma_{\min}(\boldsymbol{H}(f))}{\sigma_{\max}(\boldsymbol{H}(f))}. \tag{13}$$

When IM is close to zero, inverting $\boldsymbol{H}(z)$ is difficult, and when IM is close to 1, inverting $\boldsymbol{H}(z)$ is easy.

For the case of a symmetric stereo reproduction system with the same radius $(r_1=r_2)$, IM can be rearranged to a very simple form as follows.

$$IM(f)=\frac{\min(|C_i(f)+C_c(f)|,|C_i(f)-C_c(f)|)}{\max(|C_i(f)+C_c(f)|,|C_i(f)-C_c(f)|)}, \tag{14}$$

where, $C_i(f)$ and $C_c(f)$ are frequency response functions of $C_i(z)$ and $C_c(z)$ described in section III, respectively. It is interesting to note that in the Shuffler implementation, $IM(f)$ is equal to the minimum over the maximum gain of $A_+(f)$ and $A_-(f)$. If the two gains of the diagonalized form in the Shuffler implementation differ significantly, $IM(f)$ decreases and the inverting condition gets worse.

Figure 9 shows $IM(f)$ for a symmetric arrangement condition for symmetric angles from 0 to 150 degrees. An angle of near 0 degrees corresponds to a stereo dipole system. To provide a representative value for each case, the average value of the curve along the frequency is noted on the figure. This indicates that a symmetric stereo loudspeaker arrangement close to 70 degrees is best from the viewpoint of inverse modeling for cross-talk cancellation. This result presents a different viewpoint of the preferred stereo loudspeaker arrangement from the classical arrangement of an audio system where $\pm 30^{\circ}$ is recommended or from a stereo dipole system where a small angle less than 10 degrees is recommended.

VI. SIMULATION

The validity of the presented approach to the cross-talk canceler generation has been checked via computer simulation. A stereo loudspeaker system is considered, where the angles to the left and right loudspeakers are 10° and 30°, respectively, and the distances from the listener to the right and left loudspeakers are 2 and 2.5 meters, respectively. Though this is not a normal configuration that is recommended as a common audio system or com-

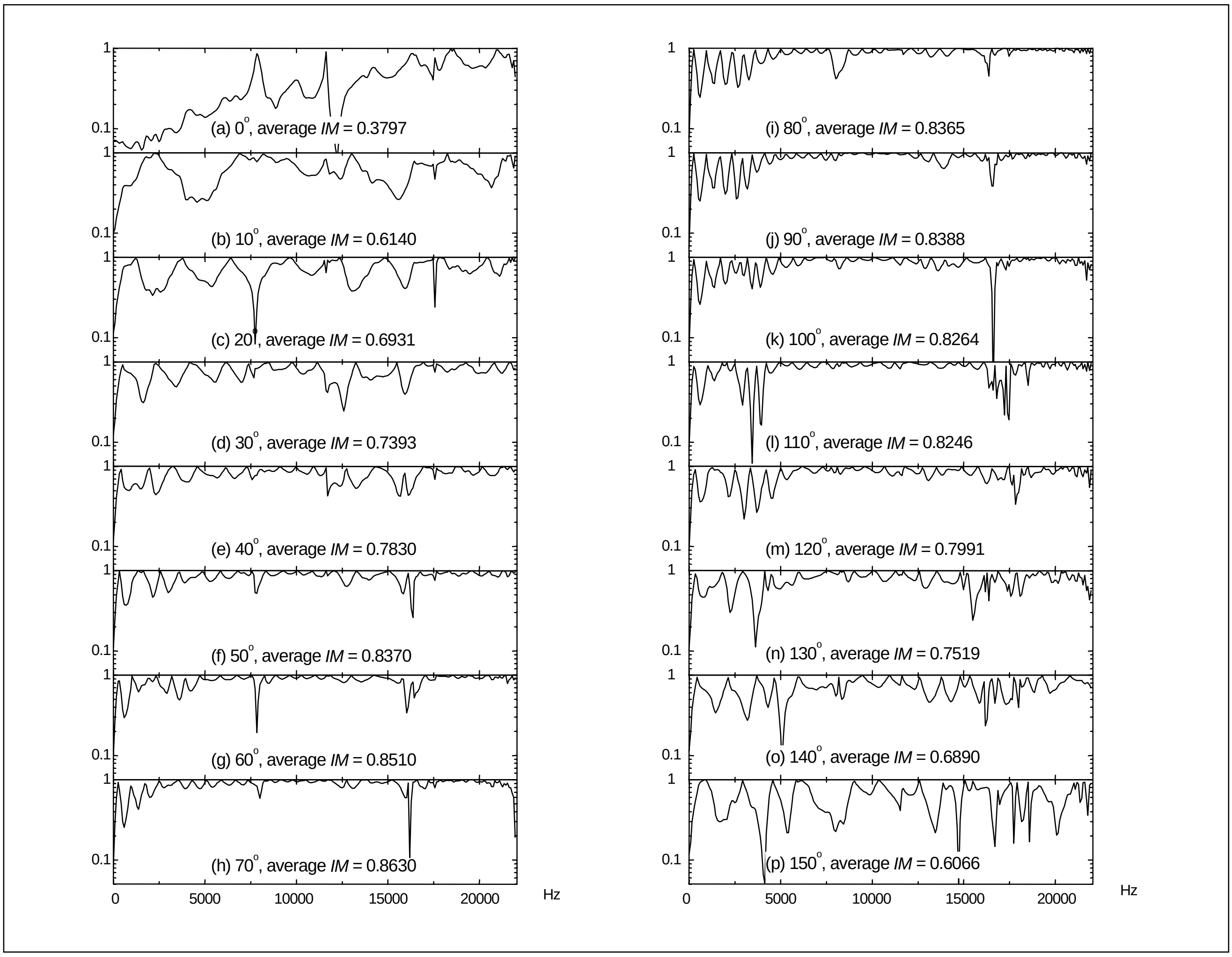

Fig. 9. Invertability measure *IM(f)* for a symmetric arrangement condition.

puter speaker system, the cross-talk canceler can reproduce the desired sound field around the listener. The sampling rate of binaural signal is assumed to be 44.1 kHz. The process of getting $\boldsymbol{C}(z)$ whose number of coefficients was 256 with $d = 128$ for a standard radius case of presented configuration was simulated—the distance to the left and right loudspeakers are 1.5 meter while the angles between them are maintained. The database of HRTFs was measured at ETRI, and 128 coefficients were used for each HRTF. The impulse response of the HRTFs consisting of $\boldsymbol{T}(z)$ and the transaural system $\boldsymbol{H}(z)$ are shown in Fig. 10(a). Four FIR-type transfer functions in $\boldsymbol{C}(z)$ were adapted using the algorithm described in Fig. 6 and they are shown in Fig. 10(b). To check the accuracy of the delayed inverse modeling, the four impulse responses consisting of $\boldsymbol{C}(z)\boldsymbol{T}(z)$ are shown in Fig. 10(c). The desired impulse response of $\boldsymbol{T}(z)\boldsymbol{C}(z) = z^{-d}\boldsymbol{I}$ is obtained well by the presented adaptation scheme. Note that the zero response of off-diagonal parts of $\boldsymbol{T}(z)\boldsymbol{C}(z)$ as in Fig. 10(c) is practically impossible because real HRTFs may differ from those used in the simulation, because the listener may move slightly around the exact prescribed position so that $\boldsymbol{H}(z)$ may change to some extent. The problem of extending the zone of equalization is another important research topic in sound reproduction with cross-talk cancellation, though it was not considered here.

VI. CONCLUSION

Building a cross-talk canceler for reproducing binaural sound with multiple loudspeakers was investigated. To simplify the building process without losing the generality, it was assumed that the direct field effect of the loudspeaker is dominant to the listener. The presented method effectively builds a cross-talk canceler for a general configuration of loudspeakers based on the cross-talk canceler database of standard-radius cases and a distance compensation process. Some efficient adaptive techniques to build the cross-talk database were presented. Also, the ar-

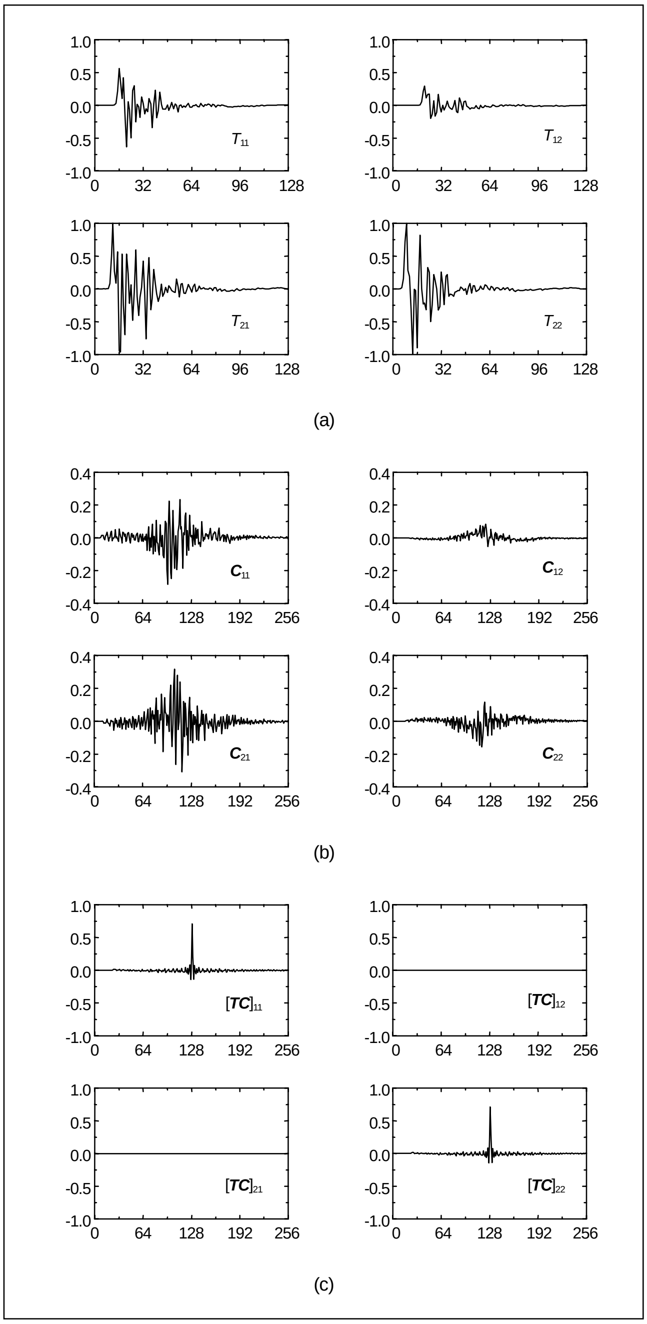

Fig. 10. The impulse response of (a) $\boldsymbol{T}(z)$, (b) $\boldsymbol{C}(z)$, and (c) $\boldsymbol{T}(z)\boldsymbol{C}(z)$ for 10° and 30° arrangement.

rangement of loudspeakers from the viewpoint of providing good inverting condition was investigated. For a symmetric stereo arrangement, the best one was found near $\pm 70^{\circ}$. To reveal the meaning inside this result needs more analysis on the HRTF. It should be noted that the invertability is not the only factor of relevance in cross-talk cancellation system. For example, the stereo dipole system provides an ill-posed inverse problem especially for low frequency, but it produces highly satisfactory results. Computer simulation results verified the usefulness of the presented approach as well as the adaptive technique to get the inverse model. As for the issues about the performance of cross-talk cancellation, it was found that caution is needed in setting up the reproduction system to avoid ill-conditioning of the sound reproduction system.

REFERENCES

[1] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*. Academic Press, Cambridge, MA, 1994.

[2] P. Damaske, "Head-Related Two-Channel Stereophony with Loudspeaker Reproduction," *J. Acoust. Soc. Am.*, Vol. 50, No. 4, 1971, pp. 1109–1115.

[3] W. G. Gardner and K. D. Martin, "HRTF Measurements of a KEMAR," J. *Acoust. Soc. Am.*, Vol. 97, No. 6, 1995, pp. 3907–3908.

[4] D. Griesinger, "Equalization and Spatial Equalization of Dummy-Head Recordings for Loudspeaker Reproduction," *J. Audio Eng. Soc.*, Vol. 37, No. 1, 2, 1989, pp. 20–29.

[5] E. M. Wenzel, M.Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindivisualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, Vol. 94, No. 1, 1993, pp. 111–123.

[6] H-S Kim *et. el.*, "Cross-talk Canceler User Interface for Two-speaker 3D Sound Reproduction," *Proceedings of Korean Society for Emotion and Sensibility*, November 1998, pp. 163–167.

[7] O. Kirkeby and P. A. Nelson, "Properties of Least Squares Inverse Filters Used for Multi-channel Sound Reproduction," *ACTIVE95*, 1995, pp. 1259–1270,

[8] P. A. Nelson, H. Hamada and S. J. Elliott, "Inverse Filters for Multi-channel Sound Reproduction," *IEEE Trans SP*, Vol. 40, No. 7, July 1992, pp. 1621–1632,

[9] S. J. Elliott, I. M. Stothers, and P. A. Nelson, "A Multiple Error LMS Algorithm and Its Application to the Active Control of Sound and Vibration," *IEEE, Trans ASSP*. Vol. 35, No. 1, 1987, pp. 1423–1434.

[10] H-S. Kim, "Active Adaptive Control for Systems with Cancellation path," Ph.D. Thesis, Dept. of Mechanical Engineering, KAIST, chapter 4, 1998.

[11] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.*, Vol. 37, No. 1, 2, 1989, pp.3–19.

Hyoun-Suk Kim was born in 1968 in Seoul, Korea. He received B.S., M.S. and Ph.D degrees in Mechanical Engineering at Korea Advanced Institute of Science and Technology (KAIST) in 1991, 1993 and 1998, respectively. He joined Electronic Telecommunications Research Institute (ETRI) in 1998, where he is currently researching audio signal processing as senior researcher at Virtual Reality Research & Development Center. He is interested in adaptive signal processing, signal processing for 3D sound reproduction, active noise/vibration control.

Poong-Min Kim received his B.S. degree in computer science from Dan-Kuk University in 1985. He received the M.S. degree in computer science from Korea University in 1993. He joined ETRI in 1985. He is a principal member of engineering staff of Auditory Information Team of Virtual Reality Department. His current interests include 3D Sound Processing of VRML.

Hyun Bin Kim received B.S. and M.S. degrees in applied statistics from Chungang University in 1985 and 1988, respectively. He received his Ph.D. degree in statistics from Okayama University in 1996. He had worked for developing S/W in System Engineering Research Institute (SERI) from 1984 to 1998. He joined ETRI in 1998, where he is currently working as team manager at the Virtual Reality Research & Development Center. His current research interests include 3-Dimentional Sound, Digital Audio Watermarking, Pattern Recognition and HCI.