A METHOD OF DESIGNING INVERSE SYSTEM FOR MULTI-CHANNEL SOUND REPRODUCTION-SYSTEM USING LEAST-NORM-SOLUTION

*A. Kaminuma, **S. Ise, *K. Shikano

*Graduate School of Information Science
Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara 630-0101, JAPAN
E-mail: atunob-k@is.aist-nara.ac.jp

**Graduate School of Engineering, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, JAPAN

1. INTRODUCTION

An inverse system for a sound reproduction must be a digital filter which is designed such that it has a stable response and a practical length. According to MINT(Multi INput output inverse Theorem)[1], an inverse system between three secondary sources and two control points can be realized as stable FIR filters. However, in order to obtain the coefficients of the FIR filters as an inverse system in a practical situation, the matrix size would be so large that in practice it would be impossible to calculate it. On the other hand, practical frequency domain design methods[2] require the number of secondary sources to be equal to the number of control points to obtain fully determined solutions. This is the main reason why transaural systems, which has two control points at the listener's ears, usually have two loudspeakers. However, because the transaural system which has two loudspeakers cannot be guaranteed to be causal and stable, the system has not been able to be realized except under limited conditions in anechoic rooms. To resolve these problems, we propose a

new design method for the multi-channel sound reproduction system using the least-norm-solution in the frequency domain. This method enables the realization of a transaural system with multiple loudspeakers by designing practical FIR filters as the inverse system, which are approximately causal and stable. In this paper, we first construct a multi-channel sound reproduction system by the proposed algorithm, and we then investigate the accuracy of the transaural system realized in real environment and evaluate its sound quality subjectively. Second, we discuss about the accuracy of the inverse system, which depends on the number of secondary sources and the number of inverse filter taps.

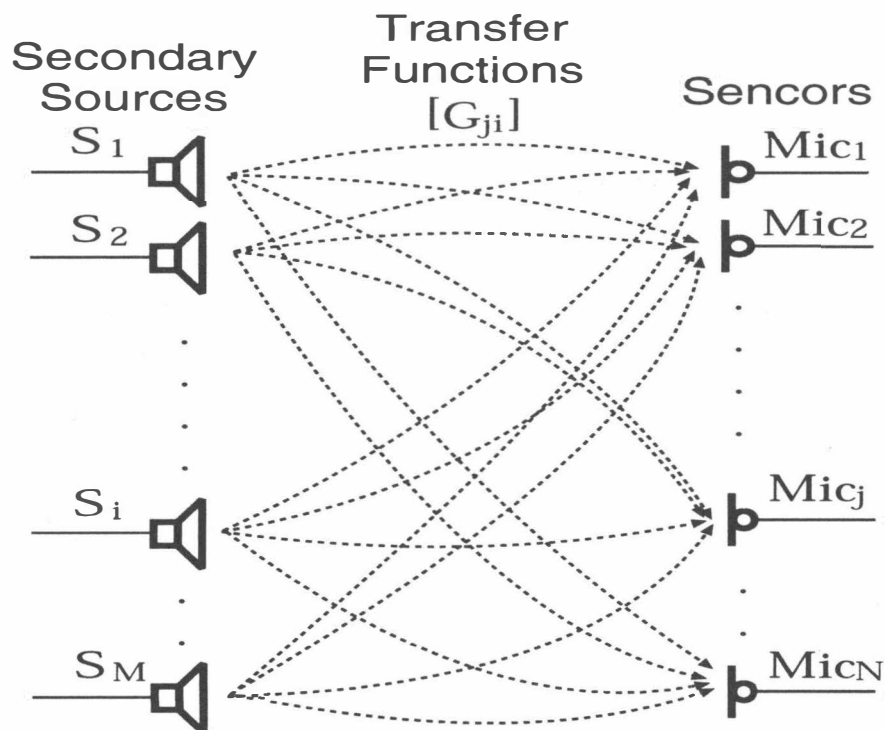## 2. INVERSE FILTER DESIGN USING THE LEAST-NORM-SOLUTION

Fig.1 Transfer functions between M secondary sources and N sencors

As illustrated in Fig.1, we assume an acoustical system $[G_{ji}]$ between a secondary source $i(i = 1 \cdots M)$ and a sensor $j(j = 1 \cdots N)$ and inverse filters $[H_{ij}]$ which totally correct the output signals of the sensors, C, to be equal to C'. The relationship between $[G_{ji}]$ and $[H_{ij}]$ is given by

$$\begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1M} \\ G_{21} & G_{22} & \cdots & G_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ G_{N1} & G_{N2} & \cdots & G_{NM} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{M1} & H_{M2} & \cdots & H_{MN} \end{bmatrix} = I_{jj} \tag{1}$$

where I_{jj} is the identity matrix. $[H_{ij}]$ can be obtained by $[H_{ij}] = [G_{ji}]^{-1}$ only if $[G_{ji}]$ is square, that is $M = N$, and the determinant of $[G_{ji}]$ is not zero. However when $M < N$, perfect correction of the acoustic system cannot be expected, $[H_{ij}]$ can be determined by the least mean square method which minimizes the error, $E[\|C - C'\|^2][3]$. When $M > N$, we can obtain the matrix solution conditionally using the pseudo-inverse of a matrix. Now consider that A is an (n, m) arbitrary oblong complex matrix, and that $\mathbf{x}$ is an m-dimension vector. We can write the following linear equation(2).

$$A\mathbf{x} = \mathbf{b} \tag{2}$$

This linear equation does not have a unique solution. Rather, it has a $n \times m$ dimension in solution space. In the case that $rank(A) = n$, we can obtain the form of the least-norm-solution generally.

$$\mathbf{x} = B(AB)^{-1}\mathbf{b} \tag{3}$$

This solution indicates the minimum of the solution space. Then we can estimate $\mathbf{x}$ by using $B = A^{\dagger}$.

$$\mathbf{x} = A^{\dagger}(AA^{\dagger})^{-1}\mathbf{b} \tag{4}$$

If A is replaced by $[G_{ji}]$, we can design the inverse filters in the frequency domain.

3. DEVELOPMENT OF A TRANSAURAL SYSTEM WITH EIGHT LOUDSPEAKERS

3.1 System

As shown in Fig.2, a transaural system with eight loudspeakers was developed. The system calculates convolution between the recorded sound and the inverse filters which was measured from room transfer functions in a reproduction room. These signals are added by each secondary source, and sent to the loudspeakers. A listener can thus perceived sound which is the same as that which would be perceive at the observation point.

3.2 Inverse filter design

We measured the impluse responses between each of the loudspeakers and each of the two points positioned about 1 cm outside the listener's ear-canals respectively. The 65536 points time stretched pulse is used in measurement, where the sampling frequency is 48 kHz, and quantized in 2^{16} levels. Furthermore, these impulse responses with 9600 points are transformed into the frequency domain using FFT. Inverse filters are calculated from the impulse responses using the least-norm-solution in equation (4). We obtain the inverse filters in the time domain using IFFT.
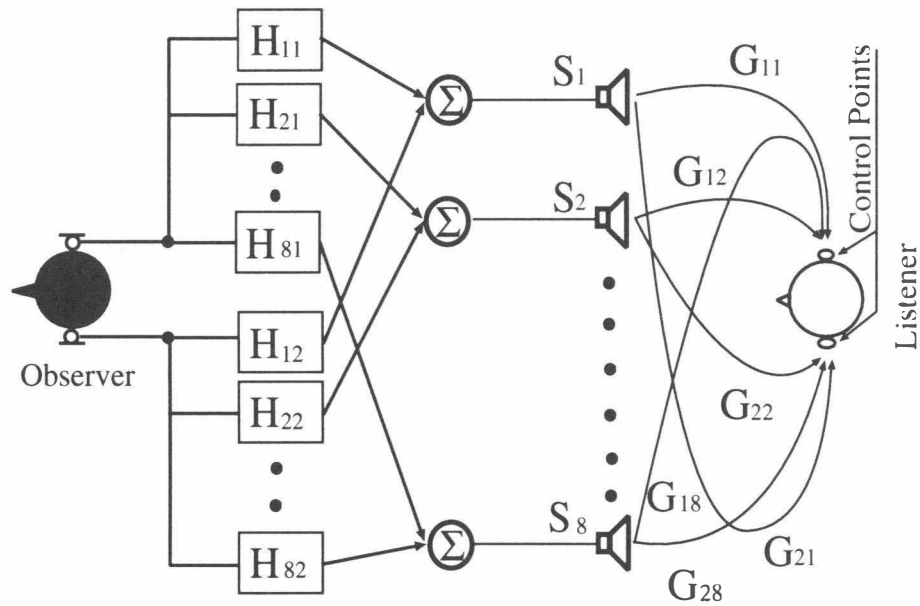
Fig.2 8-2 type sound reproduction system

3.3 Evaluating for horizontal localization

Investigating the horizontal localization of the 8-2 type sound reproduction system, we evaluated the localization subjectively through a listening test. First, we recorded the impulse response of the control position of the observer who was not included in the subject group. Inverse filters are designed using these impulse responses. We use three sources; the speech of an English male speaker, the sound of a violin, and the sound of a saxophone. These were convoluted with the inverse filters. These stimuli were filtered to remove frequencies outside the range between 10Hz and 6000Hz. The stimuli times were adjusted to 10 seconds. Twelve loudspeakers as primary sources were located in the horizontal plane around the subject at 30 degree intervals. Directly ahead of the subject was considered to be 0 degrees while directly to the right of the subject was considered to be 90 degrees. Primary sources were placed at 1.5 meter distance from the center of subject's head. Also, secondary sources were set at a distance of 2.0 meters from the center of subject's head, and all loudspeaker's height was adjusted level with the observer's ears. There were two cases of stimuli sound output; reproduction sounds from the secondary sources and primary sounds output from one of the primary sources. We subjected the subjects to twelve directions, and three kinds of sources. Each stimulus was presented twice. Subjects used head-rest to support (unfixed). The subjects had to mark the perceived position on the test papers which showed the loudspeaker positions.

Fig.4 shows the experimental results for all the subjects. The horizontal axis shows the presented direction while the vertical axis shows the perceived direction. The center of each circle represents the perceived direction. Further, the circle sizes show the frequencies perceived by all subjects. The circles are plotted against the vertical axis in 10 degree
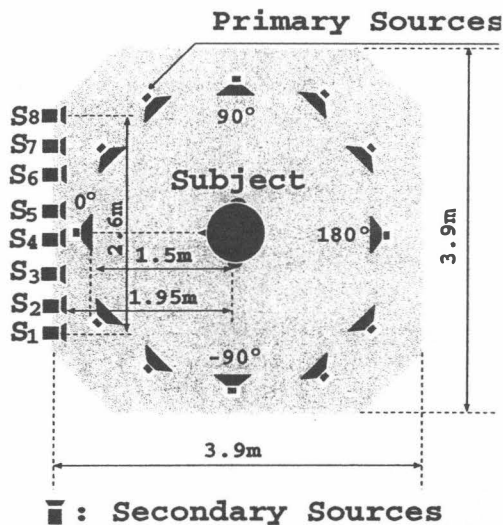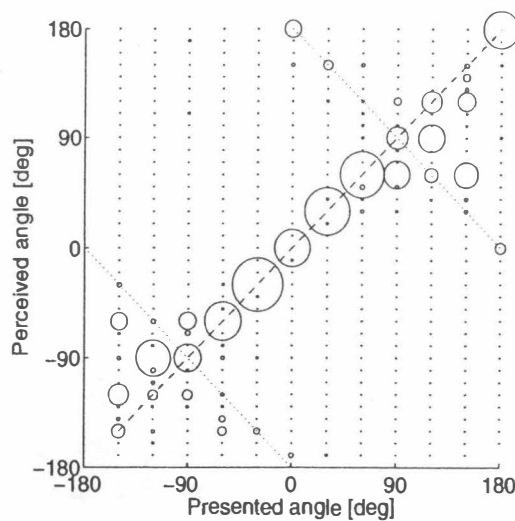
Fig.3 Experimental environments

Fig.4 Experimental results

steps. The diagonal dashed line shows the perceived direction, which agrees with the provided sound direction. The two diagonal dotted lines show the cases where the perceived direction can reflect across the median plane. Sound that were supposed to be heard from backward were sometimes perceived as comming from forward, as shown in Fig.4. The reason is considered to be the difference of HRTF between the observer which was used for measuring the transfer function and the subjects of localization test. On the whole, our system had good localization as shown in Fig.4.

4. EVALUATING SOUND QUALITY USING COMPUTER SIMULATION

In this section, we investigate the accuracy of the inverse filter for the sound reproduction system by computer simulation, especially the relationship between the number of secondary sources and inverse filter taps. This simulation uses the room transfer function recorded in a real environment.

4.1 Inverse filter design

In the previous chapter, we created an impulse response of 0.2 second and 9600 points, and transformed it into the frequency domain using FFT. We then calculated the inverse filters using the least-norm-solution(equation (4)) and by using two to eight secondary sources in steps of one. The inverse filters in the time domain was obtained using IFFT. The inverse filter length is determined by picking up 128, 256, 512, 1024, 2048, 4096 or 8192 points respectively to equalize forward and backward length from the center of filter network. The center of the filter network is selected by the average of the fastest peak and latest peak in all the inverse filter channels.
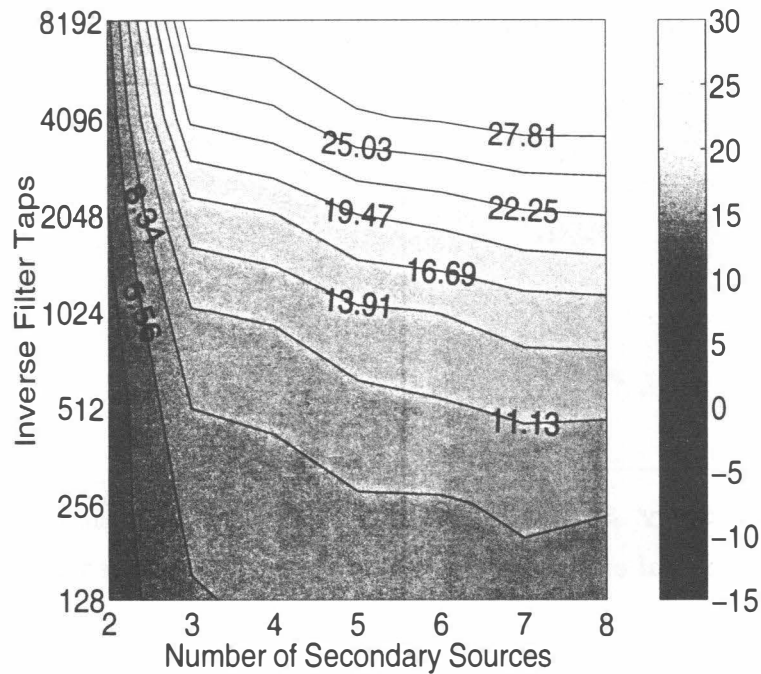
Fig.5 Experimental results with observer's transfer function

4.2 SNR (Observer's transfer function)

We assumed that no signal is arriving into the left ear of the observers. On the other hand, impulse signals are arriving into the right ear of the observers. Thus, the signals are created by the convolution of inverse filter networks and observer's transfer functions. Finally , we calculated the SNR by limiting the signals from 150Hz to 4000Hz. Fig.5 shows the result. The horizontal axis shows the number of secondary sources while the vertical axis shows the number of inverse filter taps. The number of secondary sources increase from two to eight in steps of one, and the inverse filter taps increase from 128 to 8192. The color-bar at the right in the figure is yardstick for the SNR color. It ranges from -15 to 30dB. In the figure, the numerical values on the contour lines show the SNR. Based on this figure, we easily realize that the SNR level in the case of two secondary sources is worst among any other level. For example, when the inverse filter taps is 8192, the SNR is 5 dB, which is far from 30dB of eight secondary sources's case. Due to the fact of low level SNR, those inverse filters cannot converge as a FIR filter in the case of two secondary sources and two control points, because inverse filters have common zero points. Considering over three secondary sources, large SNR is obtained when the number of secondary sources is large and the inverse filter taps number are fixed. For example, if the case of three secondary sources, SNR is 11dB. Then inverse filter taps is 1024 points. However, if the number of secondary sources is eight, SNR is 15dB. This value corresponds to the SNR case of three secondary sources, and the inverse filter taps is 2048 points. Further, in the case of three secondary sources and 1024 point inverse filter taps the same level is obtained as the case of
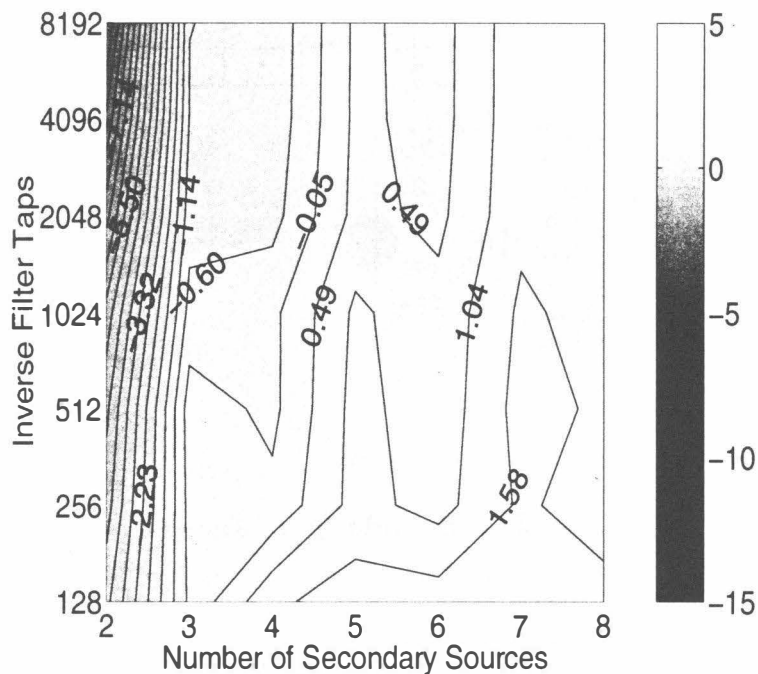
Fig.6 Experimental results with listener's transfer function

eight secondary sources and 512 inverse filter taps. By increasing the number of secondary sources to eight, smaller number of inverse filter taps is necessary in order to obtain the same accuracy than in the case of smaller number of secondary sources.

4.3 SNR (Subject's transfer functions)

We also calculated the SNR using the inverse filters introduced in the previous chapter and the transfer functions measured for five subjects. Fig.6 shows the average SNR for five subjects. The horizontal axis shows the number of secondary sources from two to eight, and vertical axis shows inverse filter taps from 128 to 8192. The color-bar at right side of the figure is yardstick depending on SNR. This scale is from -15dB to 30dB. This figure shows lower SNR comparing with the SNR in the case of using observers transfer functions. However, the SNR increases gradually as the number of secondary sources increases. More-over, in the cases of over than four secondary sources, SNR does not change so much. The increase is only about 2dB. These results suggest that when a subject listens to this repro-duction system in a real environment, the increase in the number of secondary sources leads to an improvement in accuracy. However, the auditory experiment for observers showed that for over than four secondary sources there is no improvement in accuracy. On the other hand, Fig.6 shows another interesting result. It shows that the SNR in all cases of different number of secondary sources increases with smaller number of inverse filter taps. Especially, in the case of two secondary sources, half length of inverse filter taps improves the accuracy about 1 dB. These phenomenon may be caused by increasing the transfer function error between the observer and subjects as increasing inverse filter taps.
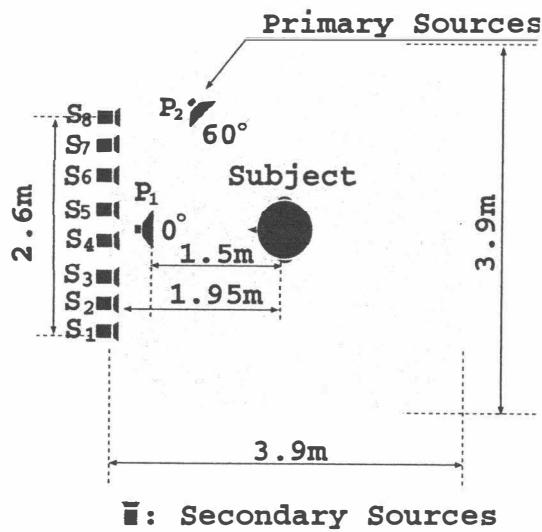
Fig.7 Experimental environments

Table.1 Arrangement of secondary sources

Secondary sources	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
2		!{					!{	
3		!{			!{		!{	
4	!{		!{			!{		!{
5	!{		!{		!{	!{		!{
6	!{	!{	!{			!{	!{	!{
7	!{	!{	!{		!{	!{	!{	!{
8	!{	!{	!{	!{	!{	!{	!{	!{

5. SOUND QUALITY TEST

Confirming the results in the section 4.2 and 4.3, we evaluate the sound quality of the proposed system using subjective experiments.

5.1 Conditions

We convolved inverse filters as seen in section 3. Two kind of stimulus; the speech of an English male speaker, and the sound of orchestra recorded in an anechoic room. The created signals were limited from 150Hz to 4000Hz using a BPF, and stimuli times were adjusted to 6 seconds. The sampling frequency is 48 kHz, and quantized in 2^{16} levels. Nine male subjects with normal hearing took part in a single half-an-hour session. Two loudspeakers with primary sources were located at 0 degree (front) and 60 degree(right) in the horizontal plane nearly 1.5 meters from the center of the subject's head as shown in Fig.7. All stimulus are presented by one of the speaker directions. Fig.7 shows the experimental environment and arrangement of the secondary sources according to Table 1. The secondary sources with outside loudspeakers were also placed at the distance of

2.0 meter from the center of subject's head. All loudspeakers were placed at the height of the observer's ears. There were two cases of stimuli sound output; reproduction sounds from the secondary sources and primary sounds output from one of the primary sources. All stimuli were presented to subjects in random order. We used seven kind of secondary sources, seven kind of inverse filter taps, two loudspeakers, and two kinds of sound sources. Each stimulus was repeated two times. Before beginning the experiments, we instructed subjects in a process of experiments by a composition and speech. Moreover, subjects used a head-rest to support, which was unfixed. The task of subjects is to write down the presented sound quality in term of existence of noise, echo and clarify. The testing papers was used from the following 7 evaluating levels in Table 2.

5.2 Results

Fig 8 shows the result of sound quality test in seven evaluating levels. The horizontal axis shows the number of secondary sources. The vertical axis shows the number of inverse filter taps. The color-bar at the right side of the figure is yardstick indicating Mean Opinion Score(MOS) ranging from one to seven points. In the figure the numerical values on contour lines mean MOS. Table 3 shows the results in the case of 2, 4, 8 secondary sources. In the case of two secondary sources, the evaluation score is low over all subject's score. According to Table 3, in the case of two secondary sources, 2.60 at maximum is showed at 256 inverse filter taps. On the other hand, the number of the inverse filter taps does not provide any information. When we listened to the reproduction sound in the case of two secondary sources, there are some noises such as burst and echo. In the case of over four secondary sources, the number of inverse filter taps do not have any effect to the evaluation score. However, it is clear that the evaluation score for four secondary sources is higher than that of two secondary sources. In the case of eight secondary sources, evaluation scores are larger than four secondary sources. However the rate of increase is less than the case of from two to four. Thus, increasing the number of secondary sources lead to an improvement of system accuracy. This effect is large until four, but is small for more than four sources. Also sound reproducted includes burst noise and other noises such as white noise. Difference between primary and secondary sources evaluation may be caused by these noise. Inverse filters including these noise at a specific frequency shows high peak compared with other frequencies. If the high peak increases, the estimation error for inverse filters also increases, such a reproduced sound includes noise.

5.3 F-test

In each category, experimental scores assigned at each levels are added, and mean and variance are calculated. Unbiased variance at each levels were analyzed using the Bertret variance test. The hypothesis that the variances were not equal was not rejected at the 95% significance level. Hence, the variance of the experimental data is uniform. Moreover we tried to test two-way layout variance analysis using the F-test. However, we could not use the test because interaction was detected. Therefore we split the data according to

Table.2 Scaling

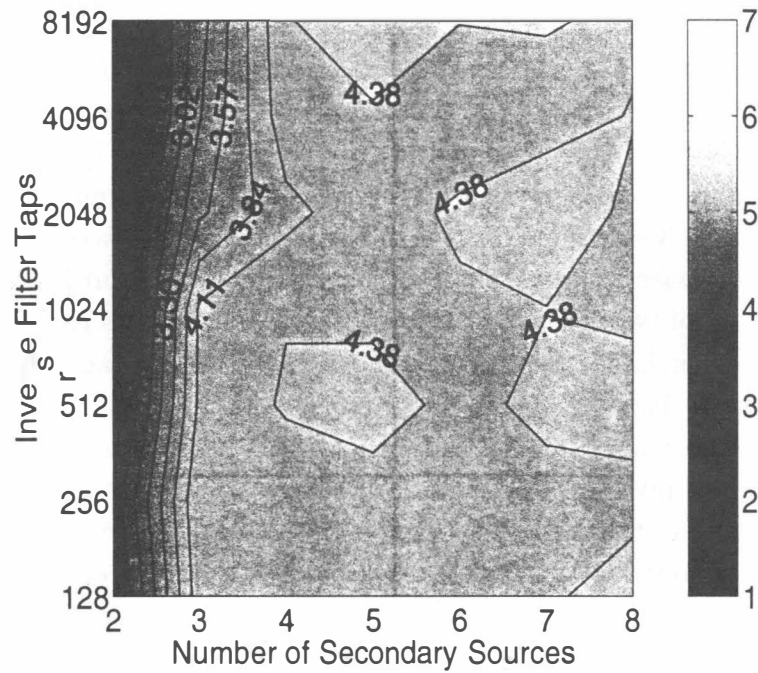|	
7.	very good
6.	good
5.	slightly good
4.	normal
3.	slightly bad
2.	bad
1.	very bad

Fig.8 Experimental results with subject's transfer function

Table.3 Experimental results with 2, 4, 8 secondary sources

inverse filter taps	2 secondary sources	4 secondary sources	8 secondary sources
8192	1.83	4.36	4.31
4096	1.67	4.28	4.40
2048	2.15	4.03	4.29
1024	1.97	4.36	4.31
512	2.19	4.41	4.54
256	2.60	4.18	4.26
128	2.37	4.29	4.57
Original		6.26	

$F_{A,B} > F_{0.05}$

8192
4096
2048

Interaction

1024
512
256
128

$F_{A,B} < F_{0.05}$

Inverse Filter Taps
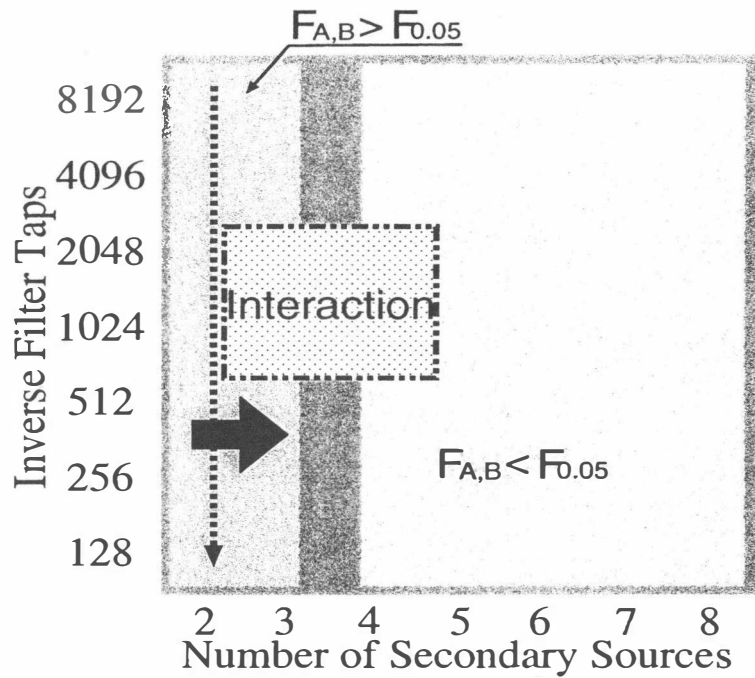
2　3　4　5　6　7　8
Number of Secondary Sources

Fig.9 F-test with subject's transfer function

secondary sources levels from two to three and from three to eight, and we tested using F-test again. Fig.9 shows the results. The horizontal axis shows the number of secondary sources, and the vertical axis shows the number of inverse filter taps. Fig 9 shows the results as follows:

- When the number of secondary sources is increased from two to three, the evaluation of the sound quality increases significantly.

- In the case of the two or three secondary sources, short inverse filter taps lead to higher evaluations of sound quality.

- When the number of secondary sources is increased from four to eight, sound quality do not change evaluations.

- In the case of the four or eight secondary sources, evaluations are not affected by the length of the inverse filter taps.

- we observe interaction on two-way layout variance analysis when the number of secondary sources is changed from three to four or inverse filter taps is changed from 512 to 2048 points.

6. CONCLUSION

We developed a transaural system with 8 loudspeakers by using a new design method of the acoustical inverse system based on the least-norm-solution. The advantages of this method are the following:

1. We can design the multi-channel inverse filter system easily, since the calculation and memory size are smaller than for a system designed in the time domain.

2. The system becomes stable as the number of secondary sources increases.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] "Inverse Filtering of Room Acoustics", M. Miyoshi and Y. Kaneda, IEEE Trans. ASSP, **36**, 2, 145–152 (1988)

[2] "Computer simulation of sound transmission in rooms", M. R. Schroeder and B. S. Atal, IEEE *Int, Conv, Rec*,**7**,150–155 (1963)

[3] P. A. Nelson and S. J. Elliott, *Active Control of Sound*, (ACADEMIC PRESS Harcourt Brace & Company, Publishers,)

[4] "An Inverse filter design for transaural system using least-norm-solution", A. Kaminuma, S. Ise, and ·K. Shikano, *J. Acoust. Soc. Japan. , Proc*,(September 1998)

[5] "A nature of an inverse filter for Multi-Channel sound reproduction system using least-norm-solution", A. Kaminuma, S. Ise, and K. Shikano, Technical Report of IEICE, EA99-13 (1999)