HRTF Measurement and Its Application for 3-D Sound Localization

*Kyeong Ok Kang, **Dong-Gyu Kang, ***Minsoo Hahn,
****Moon Jae Jho, and *****Dae-Gwon Jeong

Abstract

Based on the anthropometric data of Korean male adults, a head and torso simulator (HATS) is constructed to measure its head related transfer functions (HRTFs) which can be used for three dimensional (3-D) sound localization. The HRTFs, binaural impulse responses, are measured in an anechoic chamber using a burst maximum length sequence (MLS) signal of 65,535 samples and 32,768 samples acquisition at the sampling rate of 75.47kHz. Also measured are the impulse responses of a driving loudspeaker and some headphones for sound reproduction to get the exact HRTF of the HATS-alone. Through a post-processing procedure, the impulse-version HRTFs at the sampling frequency of 44.1 kHz, which have filter lengths of 512 points, are finally obtained.

As an application of the measured HRTFs, a 3-D sound processor for headphone reproduction has been developed. The signal intervals to be processed can be selected and each interval is manipulated to have its directionality and distance information by using corresponding HRTF and energy control.

Ⅰ. Introduction

When we hear a sound, it arrives at both of our ears and is modulated in various ways by the external auditory apparatus and the body. Until now, we have been focusing only on the temporal factors in the telecommunication research. Even though the information is nonverbal, the spatial factor is very important in communication. Therefore, the achievement of a realistic audio display is very useful for telecommunication because it provides the listener with spatial information for localization and talker identification.

In three dimensional (3-D) sound processing for a sound source, it is essential to control the corresponding spatial sound imagery. That is, it is necessary to control the perceived direction and distance of the sound image, which can be localized at desired point in three dimensional space. It has been known that the important cues for the perception of sound direction were binaural localization cues, that is, the interaural time differences (ITD) and interaural intensity differences (IID) across both ears, as a consequence of the acoustical diffraction at the body and the ear [1]. Binaural localization cues are processed at the brainstem level, and then the brain recognizes what we hear and from also the direction [2] [3]. Since we can get these cues to measure so-called head related transfer functions (HRTFs), we can localize the sound image at a spatial point corresponding to arbitrary azimuth and elevation angles, through filtering that sound source by HRTFs.

In general, however, the characteristics of these HRTFs mainly depend on the size and shape of the ear, head, shoulder, etc. Therefore, the characteristics of individual HRTFs are different from person to person. Consequently, from the macroscopic point of view, it is supposed that there may be a considerable difference in measured HRTFs between individuals, moreover between nations, especially between the Asian and the European, and that the effect of sound localization using these HRTFs may be different.

In this paper, a head and torso simulator (HATS) is constructed based on the anthropometric data of Korean male adults and the HRTFs of the HATS are measured in order to use them for 3-D sound localization. And as an application of the HRTFs, a 3-D sound processor for headphone reproduction is developed for a monophonic sound to have its directionality and distance information by using HRTFs and energy control.

* Realistic Telecommunications Section, ETRI
** Spoken Language Processing Section, ETRI
*** School of Communications Engineering, ICU
**** Acoustics and Vibration Group, KRISS
***** Department of Avionics Engineering, Hankuk Aviation University

Manuscript Received: November 1, 1997.

Ⅱ. Construction of a HATS and its HRTF measurement

2.1 A Korean standard HATS

As mentioned before, it was supposed that there would be a considerable difference in measured HRTFs between nations. In fact, from a pre-test, we could find different characteristics of HRTFs, especially over frequencies of above 6 kHz, between a dummy-head microphone according to ANSI S3.36-1985 and a person with data near the national anthropometric standards on the head and torso dimensions. So, we made a HATS based on the national anthropometric standards on the head and torso dimensions of Korean male adults, some of parameter values are shown in Table 1. Some of parameter values, which have no counterparts in the Korean standards, were used with them of ANSI S3.36-1985 or IEC Pub. 959 [4]. Measuring microphones can be positioned at the eardrum or at the ear canal entrance according to measuring purposes. The frontal and side view of the HATS is shown in Fig. 1.

Table 1. Head dimensions for the HATS.

(in millimeters)

items	mean	S.D.*	ANSI mean
head height	231**	11	125**
head breadth	163	6	152
head length	186	7	191
bitragus breadth	147	6	143

*: standard deviation
**: from head top to menton
***: from head top to the position of ear entrance

Figure 1. The frontal and side view of the HATS.

2.2 HRTF measurement of the HATS

HRTFs, the binaural impulse responses, can be considered as filters reflecting the directional filter characteristics of a human head and have been usually measured as impulse responses in time domain in order to easily produce environment context in a room via output devices such as headphones and loudspeakers.

When a sound is transmitted to the eardrum, the sound transmission path can be divided into two parts: one is the part dependent on the direction of sound incidence and the other independent on that direction. The former is the path from the sound source to the ear canal entrance and the sound transmission characteristics in this area can be given a function of the frequency and the direction of sound incidence. On the other hand, the latter, the path from the ear entrance point to the eardrum, is independent on the sound direction. Therefore, HRTFs are usually measured at the ear canal entrance point, partially because it is relatively easy to locate a measuring microphone at the ear entrance point. In our measurement, two sets of HRTFs were obtained, one for the ear canal entrance point of the left ear and the other for the eardrum point of the right ear.

The measurements were made in the anechoic chamber in KRISS (Korea Research Institute of Standards and Science). The HATS was mounted upright on a turntable which could be rotated accurately to any azimuth under the controller. The absorption rate of the anechoic chamber was beyond 99 percentage for the frequencies above 70 Hz and so we could almost ignore the effect of wall reflections on the HRTF measurements for the frequencies above 100 Hz. We can avoid interference due to room reflections by ensuring that any reflections occur well after the head response time, which is about 4.5 milliseconds. A driving loudspeaker was mounted on the circular beam of 3.5 meter diameter, made of stainless steel, which enabled accurate positioning of the loudspeaker to any elevation with respect to the HATS, and the beam was supported by a rectangular pipe of Aluminum. The beam and pipe were wholly covered with an absorption material in order to prevent sound reflections.

The measurement block diagram and a picture for the measurement are shown in Fig. 2 and Fig. 3, respectively. Mainly, the measurement system consists of one part to generate a maximum length sequence (MLS) signal and to drive a loudspeaker and another part to record the output signal of a microphone in the HATS and consequently to measure the impulse response. B&K type phantom P48 (16mm) microphone and B&K type 4134

(1/4") microphone with ear simulator were located at the left ear and right ear of the HATS, respectively. A burst MLS signal of 65,535 samples was used to drive the loudspeaker. We used Boss 101A loudspeaker whose frequency response was ±10dB in deviation over frequencies in 100 Hz~20 kHz and sound pressure level increased as frequency was increased. Such characteristics can be considered desirable because sound pressures of HRTF at high frequencies are rapidly decreased. Microphone output signal is transmitted into the MLS analyzer via a measuring amplifier and then digitized by a 12-bit A/D converter in the analyzer. The impulse response of the whole measurement system is calculated as the ratio of the input signal to the output signal in the MLS signal analyzer (as a crosscorrelation method), using 32,768 samples acquisition at the sampling rate of 75.47 kHz.

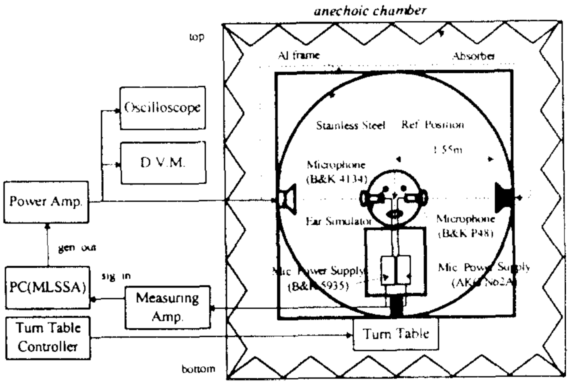


Figure 2. The measurement block diagram for the HRTFs.

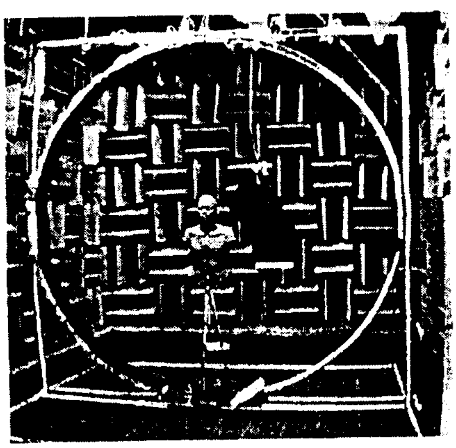


Figure 3. A picture for measuring HRTFs.

We measured the impulse responses of HRTFs at 710 points on a spherical surface of radius 1.55 meters. The angular positions of 710 points, shown in Table 2, were selected as follows. First, elevation angle was varied from -40 to 90 degrees in the step of 10 degrees. And resolution of azimuth angles was the finest at 0 degree elevation, i.e., 5 degrees, and become coarse as the elevation was changed. The measurements were made one elevation at a time, by rotating the HATS to a corresponding azimuth resolution in that elevation.

Table 2. Measurement points and azimuth resolution at each elevation(degrees in angle).

elev.	meas. pts.	△(azi.)	elev.	meas. pts.	△(azi.)	elev.	meas. pts.	△(azi.)
−40	56	6.43	10	72	5.00	60	36	10.00
−30	60	6.00	20	72	5.00	70	24	15.00
−20	72	5.00	30	60	6.00	80	12	30.00
−10	72	5.00	40	56	6.43	90	1	x
0	72	5.00	50	45	8.00			

(elev. = elevation, meas. pts. = measurement points, △(azi.) = azimuth resolution)

We also measured the impulse response of the driving loudspeaker used. We can compensate for a non-uniform speaker response by measuring the speaker response separately and creating an inverse filter. The inverse filter, when applied to an HRTF measurement, equalizes the speaker response to be flat. The impulse response of the loudspeaker was measured with the measuring microphone located at the reference position in Fig. 2, that is the midpoint of interaural axis without the HATS. At this time, the MLS signal voltage to the loudspeaker is the same as that in the measurement of the whole system. We also measured the several impulse responses of the commercial headphones to complement the frequency characteristics of the headphones themselves when the binaural reproduction was used. In the measurement for the headphones, their impulse responses were calculated as the ratio of the output signal of measuring amplifier via the measuring microphones of the HATS to that of the signal generator, with the headphones put on the HATS.

2.3 Post-processing for the HRTF DB

From the standpoint of the MLS signal analyzer, a signal sent to the generator output results in a corresponding signal appearing at the signal input. Measuring the impulse response of this system yields the impulse response of the whole system in Fig. 2.

Because the impulse responses of the loudspeaker and the headphones have no ideal characteristics, we have to compensate their responses in order to get the exact response of the HRTF for the HATS-alone. Therefore, to do so, we first transformed the measured impulse responses into the spectra in frequency domain via fast Fourier transform (FFT). Their inverse spectra were calculated, that is, the amplitudes were inversed and the phases were negated. Then we got the inverse filter for the impulse responses of the loudspeaker and headphones through inverse FFT (IFFT) on the inverse spectra. Each inverse filter will be convolved with the HRTFs measured for the whole system, to get the HRTFs for the HATS only. When the inverse filter is calculated, the original impulse response must be of minimum phase. This is because that zeroes outside the unit circle in the original filter become poles in the inverse filter.

In the following, the procedure for a post-processing to get the database of the impulse-version HRTFs used for 3-D sound process is described briefly:

(i) format conversion and normalization into 16-bit integer format,
(ii) conversion of the sampling frequencies from 75.47 kHz to 44.1 kHz,
(iii) optimization of the HRTF filter length into 512 samples long, and
(iv) implementation of the inverse filters for the loudspeaker and headphones.

Through the procedure mentioned above, we got the HRTFs and inverse filters for the loudspeaker and headphones, which have filter lengths of 512 points with minimum phase characteristics at the sampling frequency of 44.1 kHz.

Ⅲ. HRTF application for 3-D sound localization

3.1 Acoustic model of sound field

From the viewpoint of acoustical engineering, it is essential to re-establish the realistic sound field. The most feasible solution to this problem may be the application of binaural and transaural techniques, which reconstruct the desired sound field in the vicinity of each ear [5][6].

Spaciousness of a sound can be enhanced by reflected sounds reaching a listener via many sound paths indirectly, by reflecting from surfaces within the surrounding space occupied by the sound source and the listener. Among reflected sounds early reflections, especially, can give the listener impressions of the improved intelligibility and enhanced spaciousness of sound images. Moreover, it has been known that while the lateral early reflections can reinforce the spaciousness of a sound, they can reduce the localization effect of the sound.

Therefore, in an acoustic model of our 3-D sound processor, early reflections were limited to only the first order reflection which bounces once off the opposite wall relative to a sound image to be localized. The reason of this is partly because it is very difficult to describe early reflections accurately from a limited power of computations and partly because our main concerns are in sound localization using the measured HRTFs. From our point of view, this model is sufficient to solve the problem of inside-head localization (IHL) encountered in the binaural reproduction as well, considering the trade-off between spaciousness and localization of a sound.

3.2 Configuration of the 3-D sound processor and S/W structure

The 3-D sound processor was implemented as software in personal computer in C language. The block diagram and flow chart of the S/W configuration of the 3-D sound processor are given in Fig. 4 and Fig. 5, respectively. The signal intervals to be processed can be selected and each interval is manipulated to have directionality and distance information by using corresponding HRTF and energy control. A monophonic sound by microphone and/or background acoustic signals from storage media or digitized files can be inputs to be processed by the 3-D sound processor. In our model, process for a 3-D sound was made in two steps. First, localization is performed on a direct sound by a pair of HRTFs corresponding to a particular direction. And second, 1st order reflected sound opposite to the direct sound is processed by its HRTF pair to enhance a sense of spaciousness, considering time delay and energy change related to a distance. When the processed 3-D sound is played back through headphones, senses of localization and spaciousness can be perceived. And we partially solved the IHL problem by adjusting an acoustic model of reflected sound from the opposite direction.

As an application of the 3-D sound processor, it was used as one component in a realistic electronic mail [7]. The realistic electronic mail is one of useful applications for realistic telecommunication. It was implemented using key technology on stereoscopic vision, 3-D sound, and human interfaces such as speech synthesis/recognition. In this application, MPEG-compressed audiovisual 3-D signal is transmitted through a network in the form of e-mail and a receiver enjoys stereoscopic vision and 3-D sound through a MPEG decoder. The used HRTFs in this application are ones measured at the ear entrance point of the left ear of the HATS.

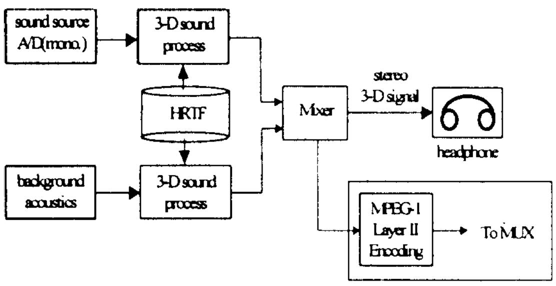


Figure 4. The 3-D sound processor.

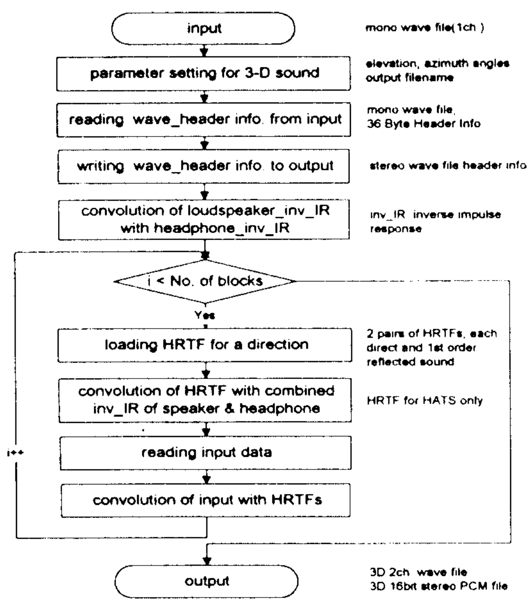


Figure 5. The S/W configuration.

Ⅳ. Conclusions

In this paper, the construction of a Korean standard HATS and the measurements of its HRTFs were described. Based on the national anthropometric standard of Korean male adults, the HATS was constructed. The HRTFs were measured at 710 points on a spherical surface of radius 1.55 meters and the impulse-version HRTFs at the sampling frequency of 44.1 kHz, which have filter lengths of 512 points, were finally obtained through a post-processing procedure.

A 3-D sound processor for headphone reproduction, which was developed as an application of the HRTFs for 3-D sound localization, was also described. In this application, the signal intervals to be processed can be selected and each interval is manipulated to have directionality and distance information by using corresponding HRTF and energy control. We partially solved the IHL problem by adjusting an acoustic model of reflected sound from the opposite direction.

Acknowledgement

This work was supported by the Ministry of Information Communication under the title of "Basic Technology in Realistic Telecommunication".

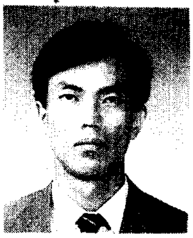
References

1. Durand R. Begault, *3-D sound for virtual reality and multimedia*, AP Professional, Chestnut Hill, MA, 1994.
2. S. H. Kang and S. H. Kim, "Realistic audio teleconferencing using binaural and auralization techniques," *ETRI journal*, vol. 18, No. 1, pp. 41-51, 1996.
3. Y. Ando, *Concert Hall Acoustics*, New York: Springer-Verlag, 1985.
4. IEC Publication 959, "Provisional head and torso simulator for acoustic measurements on air conduction hearing aids," IEC, 1990.
5. E. Wenzel, S. H. Foster, and F. Wighrman, "Realtime digital synthesis of localized auditory cues over headphones," *Proc. ICASSP '89*, 1989.
6. P. Damaske, "Head related two channel stereophony with loudspeaker reproduction," *J. Acous. Soc. Am.*, vol. 50, pp. 1109-1115, 1971.
7. K. O. Kang, D. G. Kang, and M. S. Hahn, "A 3-D sound processor for a realistic electronic mail system," presented at the 2nd Electroacoustics Conference of the Acoustical Society of Korea, Taejon, pp. 142~147, 1996, in Korean.

▲Kyeong Ok Kang

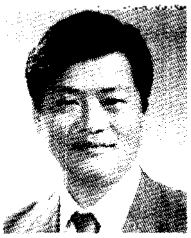
He was born in Namhae in 1962. He received B.S. and M.S. degrees both in physics at Pusan National University in 1985 and 1988, respectively, and he is now a graduate student for Ph. D. degree in electrical engineering at Hankuk Aviation University. He has joined in ETRI since 1991, and he is now a senior researcher at Realistic Telecommunications Section in ETRI. His major interests are in 3 dimensional sound and audio signal processing related to realistic telecommunications.

▲Dong-Gyu Kang



He received his BS and MS degrees in electronics engineering in 1989 and 1991 from Hoseo University, Korea, respectively. Since 1991, he has been with Electronics and Telecommunications Research Institute, as a senior researcher. His current interests are speech analysis and speech synthesis.

▲Minsoo Hahn



He was born in 1956 in Taejon, Korea. He had received BS and MS degrees in Electrical Engineering from the Seoul National University in 1979 and 1981, respectively. After three and half years research experience in the Korea Research Institute of Science and Standards, he joined the Mind-Machine Interaction Center in the Univ. of Florida as a Ph. D student in 1985. After receiving the Ph.D degree with speech-related research works from the Univ. of Florida in 1989, he joined ETRI for speech and audio research. From January 1998, he is working with the Information and Telecommunications University as a faculty member and his interested research areas include speech analysis, synthesis, and recognition, stereo and 3-D audio processing, and other signal processing topics.

▲Moon Jae Jho

His personal history was recorded in *Volume 13, Number 1E, the Journal of the Acoustical Society of Korea.*

▲Dae-Gwon Jeong

He received M.S. and Ph.D. degrees both in electrical engineering at Texas A&M University in 1987 and 1990, respectively. He had been with the Agency for Defense Development, Korea as a researcher in the guidance and control lab during 1979 to 1984. After his one-year work in ETRI in the HDTV transmission development team, he joined the faculty of Hankuk Aviation University in 1991. He is an associate professor now and works mainly on low bitrate speech and image compression including video coding. Since 1996, he has been a Rapporteur of ITU-R TG8/1 to ITU-T SG 8 for speech and multimedia coding relevant to IMT-2000 mobile systems.