where c is the sound velocity in the air (= 340 m/s). The delay times are calculated by the computer, and informed to the digital delays through MIDI. The original sound signal that simulates the environmental sound is white noise, pink noise, or true-recorded sound. The sound signal was recorded on multi track recorders (YAMAHA MT4X).

The first system can emit up to 8 reflected sounds, and then project "virtual wall." The virtual wall can change distance, direction, and width. The distance is adjusted by the delay times. The direction and width are adjusted by assignment of the reflected sound loudspeakers.

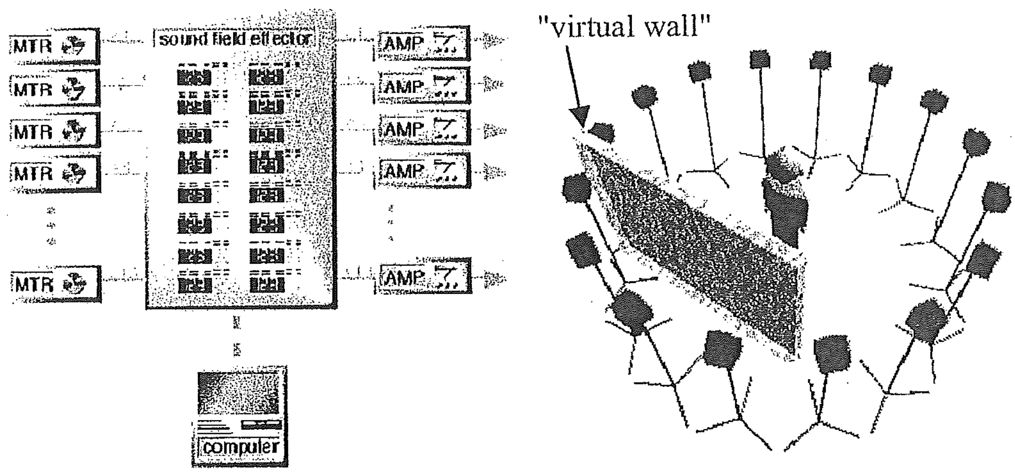


Fig. 3 Outlook of our acoustical training system for obstacle perception. This system consists of signal processor (left panel) and loudspeakers (right panel). This system can reproduce ideal sound field variation for he beginners' training. The wall shown in right panel is "virtual wall" projected by our system.

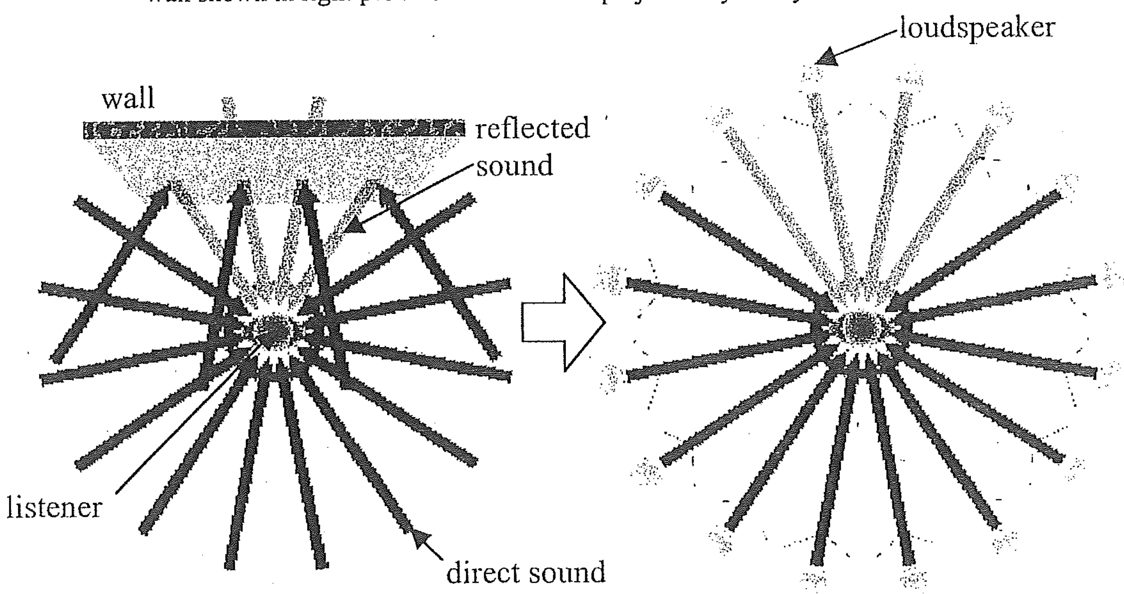


Fig. 4 Principle of reproduction of sound field variation. Our system simulates reflected sounds that reflect at surface of wall. Delay time of each reflected sound is calculated by computer, and processed by digital delays. Left panel shows environmental sound field where some sounds reflect at wall, and right panel shows simulated sound field by our loudspeaker system.

Many O&M instructors and the people concerned with blind education or rehabilitation have tried this system, and they experienced that our system can reproduce the ideal sound variation that is easier to understand for beginners than conventional training.

However, our first system was too expensive and too large to be introduced into the facilities for blind education or rehabilitation. So we also developed the simplified version that consists of 1 or 2 digital delays and 2 or 4 loudspeakers. The simplified version is now working in the Collage of National Rehabilitation Center for the Disabled, Japan for educational use.

2. Training CD

Our training system, even if it is simplified, requires economic load and space to set its hardware, when it is introduced into the blind facilities. In order to reduce economic load and save space of the blind facilities, we developed the training audio CD that contains ideal sound field data in 2001 - 2002, and are now distributing it free of charge.

Our training CD can be reproduced by general audio equipment for home use, and it can also generate the ideal sound field variation as well as the system. Our training CD contains one reflected sound in the right channel and one direct sound in the left channel. It is just the same as the simplified system of 1 delay and 2 loudspeakers.

When play this CD, the two stereophonic loudspeakers are arranged 2.4 - 3.0 m apart and facing each other. The listener's head is at the center of the two loudspeakers. The "virtual wall" appears in the direction of the right channel loudspeaker. (Fig. 5 right)

The sound field data were calculated by sound processing software (developed by REAL Software REALbasic) in computer (Apple PowerMac G4), and are recorded by CD burner (Apple iTunes & SuperDrive).

We first developed the trial version of the training CD, "Version 0.0" in 2001 and distributed it to over 50 facilities concerned with blind education or rehabilitation in Japan. Then we modified all data by their opinions, and developed the next version "Version 1.0" (Fig.5 left) in 2002.

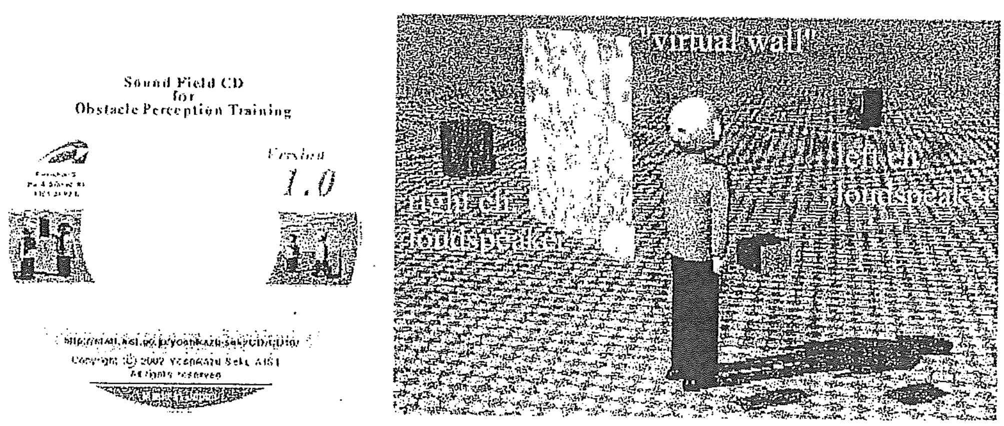


Fig. 5 Outlook of our training CD "Sound Field CD for Obstacle Perception Training, Version 1.0," and schematic explanation of reproduction of sound field. This CD can be reproduced by home use audio equipment. Two stereophonic loudspeakers are arranged 2.4 - 3.0 m apart and facing each other. Listener's head is at center of two loudspeakers. "Virtual wall" appears in direction of right channel loudspeaker.

"Version 1.0" is focused on the usefulness in practical training scene, and contains various situations. The original sound signal that simulates the environmental sound is white noise, or true-recorded sound. The sound data simulate the wall movement in various ways. For example, wall approaches, approaches and goes away, jumps, and moves randomly (Fig. 6). Total of 21 movement patterns are simulated.

Fig. 6 Examples of the simulated wall movement in the training CD "Version 1.0." Horizontal axis is time in second, and vertical axis is distance in meter. Left panel shows the approaching movement of the wall ("v" means constant velocity), and right panel shows one example of the random movements. Total of 21 movement patterns are simulated in "Version 1.0."

CD of "Version 1.0" is a single size (8 cm) audio CD, and contains 42 tracks data. Total recording time is 21 minutes 49 seconds. We are now distribution it to the facilities of blind education or rehabilitation free of charge. We are also providing the free download site of the sound field data in WAVE file format on internet. The information about the training CD, the WAVE files, and the training manual are available on our web site (URL is shown in **Contact**).

3. Conclusion

In this paper, we reported our studies to develop new acoustical training system or training CD for acquiring obstacle perception. Our artificial sound field technology can provide ideal training environment for the beginners of obstacle perception. We hope that this technology can be of help to the blind education or rehabilitation.

Our training CD "Version 1.0" has been distributed to over 40 blind facilities in Japan. We are now trying to distribute it to abroad. If wish to try it, do not hesitate to contact the following address.

Contact

Yoshikazu SEKI,
AIST,
Tsukuba 305-8566 Japan.
E-mail : yoshikazu-seki@aist.go.jp
URL : http://staff.aist.go.jp/yoshikazu-seki/CD/CD10/

References

[1] B. B. Blasch, W. R. Wiener, and R. L. Welsh, Editors, Foundations of Orientation and Mobility Second Edition, AFB Press, New York, 1997.

[2] Y. Seki, "Systematic Auditory Training of Obstacle Sense for the Visually Impaired by using Acoustical VR System," Human-Computer Interaction (Proceedings of HCI International '99 Munich) Vol. 2, 999-1003, 1999.

Coloration Perception Depending on Sound Direction

Yoshikazu Seki and Kiyohide Ito

Abstract—**Coloration is a phenomenon in which timbre changes when reflected and direct sounds are mixed. We studied the relationship between the perception of coloration and direction for two sounds. Our psychological experiments using 11 subjects suggested that a 50% threshold of coloration appears to have no difference depending on direction. When the level ratio of two sounds is closer to 0 dB, a difference appears: If direct sound comes from a lateral direction and reflected sound comes from the opposite direction, coloration perception does not increase monotonically even if the ratio approaches 0 dB. We assumed that the difference depending on direction resulted from the directional dependence of the spectrum including the head-related transfer function (HRTF) and proposed a numerical model for predicting psychological results using the comb structure on the spectrum observed at the eardrum. We measured spectra using a head and torso simulator (HATS) and calculated the area, eventually finding a quantitative relationship between the area and psychological results and proposing a prediction model based on this relationship.**

Index Terms—**Acoustic reflection, architectural acoustics, auditory system, coloration, pitch perception.**

I. INTRODUCTION

WHEN direct and reflected sounds are mixed, a listener often perceives that sound timbre changes. Mixing two sounds leads to phase interference, and comb structures having dips and peaks at even intervals appear on the power spectrum contour. Assuming that the delay of reflected sound to direct sound is T, peak frequencies of the comb structure are $n/T(n = 0, 1, 2, \ldots)$ and dip frequencies are $(n + 0.5)/T$. As the power difference between direct and reflected sound decreases, the depth from dip to peak increases. The comb structure changes timbre, at which time a listener often perceives a change in pitch. The phenomenon in which a listener perceives a timbre change due to phase interference is called "coloration" (e.g., [1]), and perceived pitch is called "repetition pitch" (e.g., [2]). Coloration and repetition pitch have interested researchers studying room acoustics, audio technology, etc. In room acoustics, for example, coloration caused by the reflection from boundary walls may influence perceived sound quality and speech intelligibility [3]. In contrast, repetition pitch caused by reflected sound may be a clue that blind people use to perceive the distance to a wall *not visible* to them [4].

Much work on repetition pitch has resulted in the proposal of models to predict its performance. Thurlow *et al.* [5], [6] studied repetition pitch perception for periodic stimuli and pulse trains, concluding that pitch correlates with the time separation of two pulses. Bilsen *et al.* [2] detailed the relationship between pitch and parameters such as delay and sensation level, pointing out that coloration detection and repetition pitch discrimination are different tasks because pitch sensation degrades rapidly for delays shorter than 1.25 ms or longer than 20 ms, while coloration due to spectral energy distribution remains unchanged. Yost *et al.* [7], [8] and Yost [9] studied the relationship between strength of pitch and "ripple noise" produced by adding (subtracting) a delayed version of noise back to (from) undelayed noise. They reported that perceived pitch is strongest when ripple noise energy is contained in a spectral region centered at $4/T$. This region is the same as that described as the dominant region. They proposed a model that predicts repetition pitch and pitch strength associated with ripple noise.

In other work related to repetition pitch, Warren *et al.* [10] reported that, while repetition pitch cannot be heard for delay longer than about 20 ms, echo delay can be perceived to 500 ms. Auditory temporal analysis thus permits detection of echo repetition well below the pitch limit of $T = 20$ ms. Yost *et al.* [11] reported on a discrimination task for a complex spectral profile with temporal change. Spacing or location of spectral peaks of ripple noise changes sinusoidally as a function of time, and a listener discriminates between this stimulus and flat-spectrum stationary noise. Their results suggest that ripple noise with temporally varying peaks was indiscriminable from flat noise beyond a temporal modulation of 5–10 cycles/s. Warren *et al.* [12] detailed mechanisms of repetition pitch associated with cophasic mixtures and polarity inverted or antiphasic mixtures. Pierce [13] discussed the relationship between periodicity and the pitch perception mechanism using repeated bursts of sinusoidal tones.

Much work on coloration has resulted in the proposal of models for predicting coloration and repetition pitch performance, but models or theories of repetition pitch cannot be applied directly to explain coloration performance because as Bilsen *et al.* [2] suggested, detection of coloration and repetition pitch discrimination are different tasks. In many cases, repetition pitch has been studied monaurally or diotically, whereas coloration in actual situations should be discussed related to listening in a sound field such as a concert hall or audio listening room. Theories of coloration should therefore be constructed originally, if possible, in binaural or dichotic listening environments.

Ando *et al.* [1] studied the perception of coloration in which loudspeakers for direct and reflected sound are arranged frontally to a listener in a free field. They assumed that autocorrelation analysis occurs in the neural part of the auditory system within the time domain, and proposed a model to explain

Manuscript received July 15, 2002; revised July 3, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Walter Kellermann.

Y. Seki is with the Institute for Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology, Ibaraki 305-8566, Japan (e-mail: yoshikazu-seki@aist.go.jp).

K. Ito is with the Future University-Hakodate, Hokkaido 041-8655, Japan (e-mail: itokiyo@fun.ac.jp).

Digital Object Identifier 10.1109/TSA.2003.818032

the perception of coloration using autocorrelation. Kates [3] proposed a more advanced model explaining the perception of coloration in filtered Gaussian noise using a central spectrum display in an auditory system. Their model accurately predicted the perception of coloration obtained by diotic experiments. Meynail *et al.* [14] proposed objectively measuring coloration caused by reverberation using impulse response captured in a room. According to their method, the distribution of the mathematically processed frequency response is analyzed and standard deviation calculated. If the processed frequency response is a theoretical Rayleigh distributed one, the standard deviation is 0.523. A larger value is obtained if there is coloration. They showed that the correlation coefficient between subjective evaluation of coloration and objective standard deviation is very large—0.91. Boussard *et al.* [15] studied the perception of coloration by lateral echo. Much conventional work concerned coloration in which both direct and reflected sound came from the same frontal direction. They conducted experiments to test coloration with reflected sound from a lateral direction (azimuth from 0° to 120°). Their initial results suggested that coloration thresholds with lateral echo are much lower than those with frontal echo. Coloration thresholds with lateral echo appear to result from two cues: spectral and spatial. As they reduce the spatial cue, thresholds in both directions do not appear to differ. They concluded that coloration perception does not depend on the echo azimuth.

We report the relationship between the perception of coloration and directions of direct and reflected sound. Conventional work has mentioned coloration only in a sound field where direct sound comes from a frontal direction (azimuth $= 0°$) and reflected sound comes from the same or another direction. If direct sound comes, however, from a *nonfrontal* direction and reflected sound comes from another direction, is the perception of coloration the same as that from direct frontal sound? To find out, we conducted psychological experiments comparing perceptions of coloration in sound fields where direct and reflected sounds come from various directions. We propose a numerical model explaining results of the perception of coloration depending on direction and other parameters using objective spectrum analysis.

II. Psychological Experiments

A. Methods

1) Subjects: Subjects were 11 listeners—six males and five females averaging 24.82 years of age (SD $= 5.06$)—with normal hearing. They were not musically trained and had no previous experience in listening experiments.

2) Experimental Setup: Experiments were conducted in an anechoic room (Fig. 1). The subjects were blindfolded and seated on a chair in the center of the room. Four loudspeakers (BOSE 111AD) were arranged in four directions: front (F), back (B), left (L), and right (R) on a horizontal plane centering on the subject's head. Loudspeakers and the center of the head were 1.8 m apart. The subject held the switch box for responses.

A stimulus signal generated by a noise generator (Rion SF-05) was split into two lines, one going through a digital delay element (Roland SDE-330). The digitally delayed signal

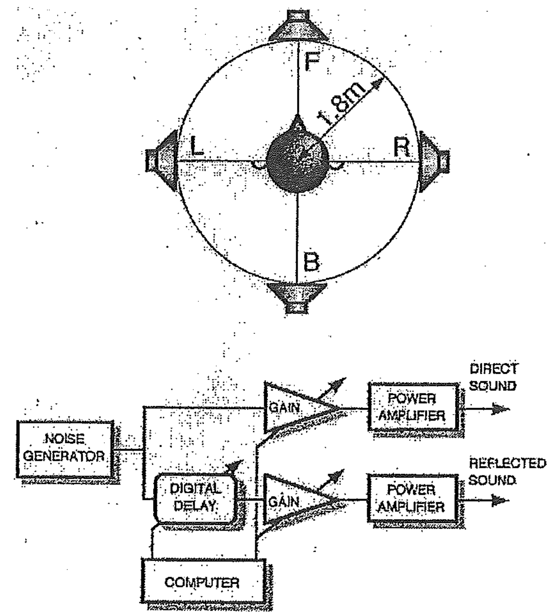


Fig. 1. Experimental setup for the psychological measurement of coloration. Above is the arrangement of loudspeakers and a subject in the anechoic room. Two of the four loudspeakers are used as a direct and reflected sound sources. Below is the signal flow. The signal of the stimulus sound is generated by a noise generator, and the delay in reflected sound is controlled by a digital delay element.

was used as reflected sound and the nondelayed signal as direct sound (in the following explanations, we use the term of "reflected sound" as the sound of the delayed signal and "direct sound" as nondelayed). Signals were gain-controlled (built-in Roland SDE-330 gain controller) and transferred to two of four loudspeakers through amplifiers (BOSE 1702). A computer (NEC PC-98Ap3) calculated delay and gain, controlling the digital delay element and gain controller through the musical instrument digital interface (MIDI).

3) Stimuli: Stationary pink noise in a 40 Hz–16 kHz frequency band was used as the stimulus signal. Total sound pressure of reflected and direct sound was 63 dB (A) at the center of the subject's head. The sound pressure was measured by a sound pressure level meter (B&K 2233) at the position of the subject's head, with the subject's head not in place.

4) Procedures: Selecting pairs of two directions for direct and reflected sound from F, B, L, and R yields seven pairs: FB, FR, BF, BR, RF, RB, and RL, where the first letter indicates the direction of direct sound and the second indicates reflected sound. Pairs FL, BL, LF, LB, and LR are omitted because they are symmetrical for FR, BR, RF, RB, and RL. These seven pairs were used as a condition of directions in experiments. In each experiment, only two loudspeakers selected as the pair of directions were used and the other two were removed from the anechoic room.

The delay of reflected sound was 2 ms, and the lowest peak frequency of the comb structure was 500 Hz.

Level ratios of reflected and direct sound were from 0 dB to -22.5 dB at -1.5 dB steps, with 16 ratios in all. Total sound pressure of reflected and direct sound was kept at 63 dB(A) at the center of the subject's head in any ratio by adjusting gain controllers.

In the timing chart of stimulation and response (Fig. 2), the comparison stimulus consisting of direct and reflected sound was emitted through the selected pair of loudspeakers. In the standard stimulus, reflected sound was replaced by pink noise noncorrelated with direct sound and having the same spectrum, sound pressure, and stochastic distribution as reflected sound in the comparison stimulus. If coloration occurs in the comparison stimulus, a listener can distinguish timbres of the two stimuli, because the standard stimulus is not colored under any condition even if the comparison stimulus is colored. If coloration does not occur, timbres of the two stimuli are not distinguishable. The stimulus lasted 1 s, including a 0.3 s rise and a 0.3 s decay. A subject listened to two stimuli in a sequence of standard and comparison, and responded whether timbres of the two stimuli were distinguishable by pressing a button on the switch box within 1 s of stimulation.

Subjects took part in 20 trials $\times$ 16 ratios $\times$ 7 pairs of directions, totaling 2240 trials per subject. Trials were divided into 28 sessions. One session consisted of 80 trials (=1 pair $\times$ 16 ratios $\times$ 5 trials) arranged randomly, and took 4 min (=80 trials $\times$ 3 s). These sessions were tested randomly. The subject rested for 4 min at each interval between sessions. During the rest, the experimenter asked the subject whether he/she was tired or not, and checked whether the subject could continue the experiment. Fortunately, no subject showed symptoms of fatigue through the experiment. Total time of experiments for one subject was 220 min (=28 sessions $\times$ 4 min + 27 intervals $\times$ 4 min).

A subject was given examples of direct and reflected sounds preceding each session. Generally, if the delay of reflected sound was less than 30–50 ms (this value depends on the level ratio of direct sound and reflected sound), listeners perceived a sound image in the direction of direct sound but not that of reflected sound because of the precedence effect (cf. [16]). Listeners may thus perceive clear sound images in both directions when the standard stimulus is presented, but may perceive a sound image only in the direction of direct sound when the comparison stimulus is presented. The subjects were told preceding experiments that they may perceive a change in sound image but were not to respond to the timbre change, and were also instructed to compare timbres of sounds in the direction of direct sound, not compare reflected sound images.

The instruction to the subjects was as follows: "Listen to the standard stimulus and comparison stimulus, and compare timbre of sound in the direction of the loudspeaker of direct sound. If you perceive the timbre change, please respond by pressing the button. In some case, you may hear sounds from the other directions, but this sound is not a matter. You must pay attention only to the direction of the direct sound loudspeaker."

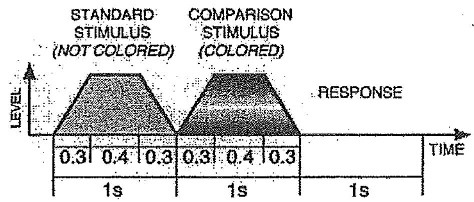


Fig. 2. Timing chart of stimulation and response. The horizontal axis indicates time and the vertical axis level. The comparison stimulus consists of direct and reflected sounds. In the standard stimulus, reflected sound is replaced by noncorrelated pink noise. Subjects compare the two stimuli, responding whether their timbres are distinguishable within 1s.

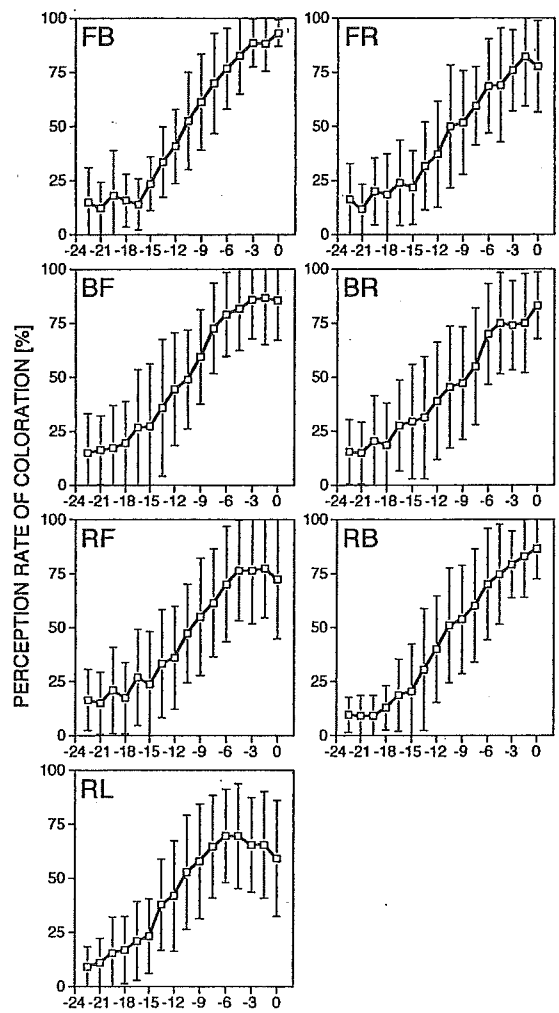


Fig. 3. Perception of coloration as a function of level ratio of reflected sound versus direct sound. The seven panels correspond to the seven pairs of directions. The horizontal axis indicates the sound pressure level ratio of reflected versus direct sound. The vertical axis shows the total percentage of 11 subjects' responses. The error bar shows standard deviation among subjects.

B. Results

In results of psychological experiments (Fig. 3), perceptions of coloration are total percentages of 11 subjects' responses.

Variations in perception for the ratio are statistically significant for all seven pairs of directions (ANOVA, $p \ll 0.01$). For ratios below about -6 dB, perceptions of coloration increase almost monotonically as the ratio increases, and differences among pairs of directions are not found. The 50% threshold of the perception of coloration is at about -10 dB in all seven pairs, and does not depend on direction. Our results agree with those of Boussard *et al.* [15].

At the ratio above -6 dB, some differences occur among pairs, particularly RL, which shows a marked decrease as the ratio increases above -4.5 dB. The perception at 0 dB is also significantly lower than that of the other six pairs (t-test, $p < 0.01$ for FB, BF, BR, and RB; $p < 0.05$ for FR, and RF), which we considered due to the diffraction loss by the head. For pair

RL, both direct and reflected sounds come from a lateral direction, so both ears must observe one of two sounds as a diffracted sound that comes half way around the head. For the left ear, the reflected sound coming from the left arrives directly, but direct sound coming from the right must be diffracted half way around the head to arrive and is weakened. For the right ear, direct sound arrives directly, and the reflected sound diffracts and is weakened. Finally, because one of the two phase-interfering sounds is weakened by diffraction loss, comb structures are reduced at both ears, which may be why the result of RL at around 0 dB is lower than the others. The peak at about -4.5 dB in RL means that, at the left ear, the power of reflected sound attenuated about -4.5 dB is nearest to the power of direct sound weakened by diffraction. Slight differences occurring among the six pairs except for RL at ratios above -6 dB are considered due to a difference of diffraction or reflection aspects of the ear pinna depending on direction.

III. Numerical Model

A. Principle

The reason for the difference among pairs was discussed based on sound propagation such as diffraction in Section II-B to explain the difference in perception of coloration depending qualitatively on directions. Can we explain this, however, more quantitatively?

Coloration is perceived by the comb structure on the spectrum. The spectrum observed at the eardrum varies depending on direction because it includes the strongly directionally dependent head-related transfer function (HRTF). We thus consider that the difference in perception depending on direction results from differences in comb structures on the spectrum observed at the eardrum depend on direction. We could construct a model explaining the difference in perception depending on direction using spectra including HRTF. As stated in the Introduction, some conventional work has proposed models to explain the perception of coloration in free-field listening using objectively measured materials, such as the model of autocorrelation function [1], and the method of impulse response analysis [14]. These models, however, omit the difference between the sound source signal and observed signal at the eardrum influenced by HRTF and thus do not account for the difference depending on sound direction. We proposed a model using spectra observed at the eardrum.

Consider the relationship between the observed spectrum and the perception of coloration in our experiments (Fig. 4) using samples of spectrum measured at the eardrum. These spectra were measured at the eardrum of the right ear using the head and torso simulator (HATS) (B&K 4128) located at the same position as the subject of the psychological experiment. Fast Fourier transform (FFT) analysis was conducted at 50-kHz 4096-point sampling using the directional pair FB and ratios of 0, -12, and -22.5 dB and measuring both standard and comparison stimuli. Note the markedly large comb structure with peaks at 500 Hz and multiples in the spectrum of the comparison stimulus at 0 dB, compared to the smooth contour of the standard stimulus. The comb structure of the comparison stimulus appears to decrease as the ratio decreases. In the difference of the spectrum given by subtracting the spectrum of the standard stimulus

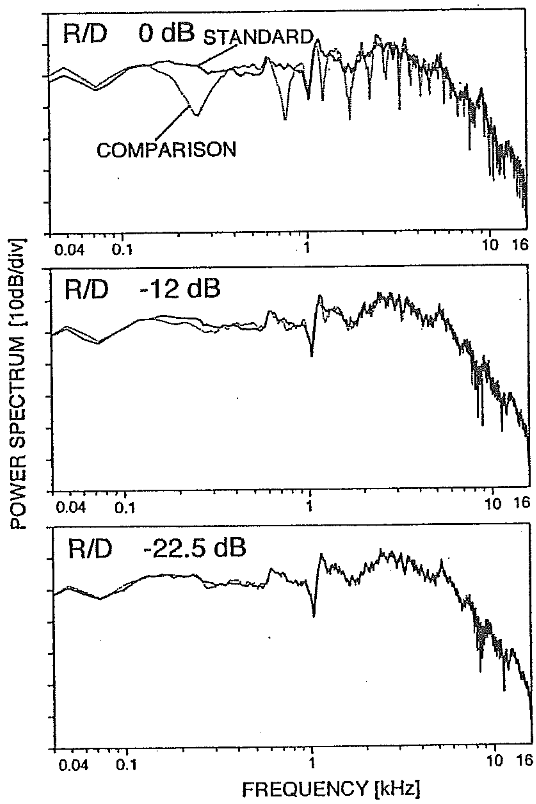


Fig. 4. Examples of spectra observed at the right eardrum. The pair of directions is FB. Above is a ratio of 0, at center a ratio of -12, and below a ratio of -22.5 dB. The horizontal axis indicates frequency and the vertical axis the relative level of the power spectrum in 10 dB/div. The solid line shows the standard stimulus and the dotted line the comparison.

from that of the comparison (Fig. 5) for a ratio of 0 dB, the comb structure appears widely across frequencies and has large depth. Subjects perceived coloration strongly at about 93% for this ratio. In contrast, the depth of the spectrum at -12 dB is smaller except at 9 kHz. For -22.5 dB, little ripple is seen in the spectrum. The perceptions of coloration at -12 and -22.5 dB are about 41% and 15%. From these examples, we expect that the magnitude of the comb structure is at least monotonically related to the perception of coloration.

As a basic concept of our model, we hypothesize that the *area* of comb structures on the spectrum at the eardrum dominates the perception of coloration. The above examples suggested that, when the bandwidth of the signal is wide enough to contain peaks and dips, the perception of coloration becomes stronger as the depth from the peak to the dip deepens and the frequency bandwidth of the comb structure (i.e., the width of the frequency range where the comb structure appears on the spectrum) widens. The area given by multiplying depth and bandwidth can thus be used as a factor to explain the perception of coloration. In the definition of the area of the comb structure (Fig. 6), that between the upper and lower envelopes is used as a factor to explain the perception of coloration in our model. The scale of frequency is logarithmic to correspond to the auditory scale. If we apply this definition to examples in Fig. 5 and calculate the areas, we observe the magnitude of the comb structure more objectively.

Consider the actual method used to find the area from observed data. If we try to find the area directly from the measured

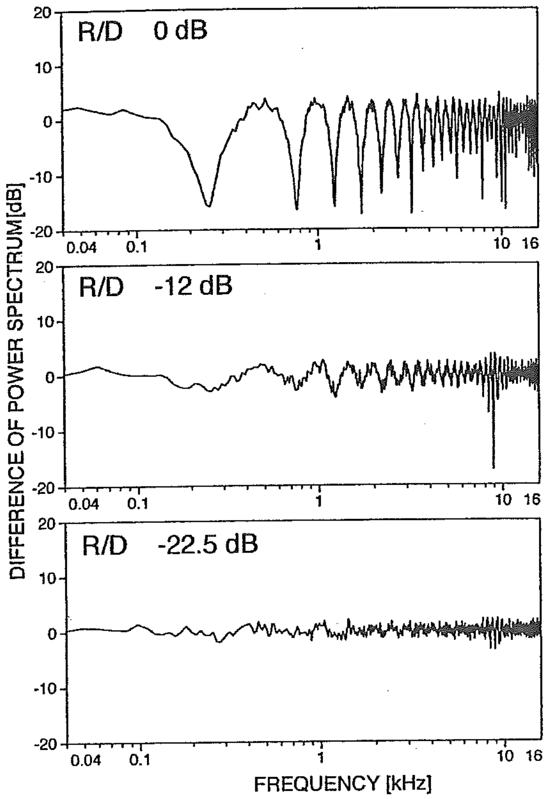


Fig. 5. Difference in power spectrum found by subtracting the spectrum of the standard stimulus from that of the comparison (Fig. 4). Above is a ratio of 0, at center a ratio of -12, and below a ratio of -22.5 dB. The horizontal axis indicates frequency and the vertical axis the level difference of the power spectrum.

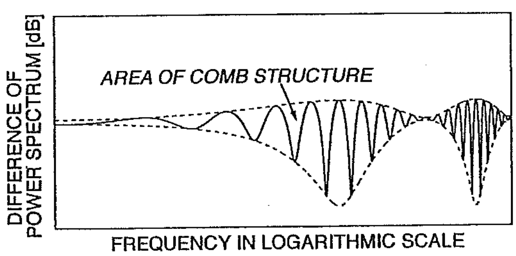


Fig. 6. Area of the comb structure used to explain the perception of coloration in our model. The horizontal axis indicates frequency in logarithmic scale and the vertical scale the level difference in dB. The solid line shows the difference of the power spectrum found by subtracting the spectrum of the standard stimulus from that of the comparison. The dashed line shows the envelope of the comb structure. The area of the comb structure is the magnitude of the domain between upper and lower envelopes.

power spectrum difference (Fig. 5), we must draw the envelopes by connecting the tips of peaks or dips using freehand or approximate curves that involves large error because we cannot precisely know the location of the tip of the dip from the measured spectrum (Fig. 5) if actual dip is very deep and frequency resolution of the FFT analysis is not high enough to show the accurate frequency of the tip. The depth of dips in a frequency range less than about 2 kHz at 0 dB (at left in upper Fig. 5), for example, is only about 15–20 dB, while the accurate depth of dips is estimated to be much deeper because the power difference of reflected and direct sounds observed at the eardrum must be very small at such a low frequency at 0 dB and must give rise to stronger phase interference. We thus cannot find the area directly from the measured power spectrum difference, requiring another method.

The algorithm we propose is as follows: Assuming that the power difference of reflected and direct sounds observed at the eardrum is $A(f)$ dB as a function of frequency f, the difference between the upper and lower envelopes of the comb structure $G(f)$ dB is found by (1)

$$G(f) = 20\log_{10}\left|\frac{1+10^{A(f)/20}}{1-10^{A(f)/20}}\right|. \tag{1}$$

We define that the spectrum of direct and reflected sounds observed at the eardrum as $D(\omega)$, and $R(\omega)$, where ω is the angular frequency. Assuming that the directions of direct and reflected sounds are Θ and Φ, $D(\omega)$ and $R(\omega)$ are described as in (2)

$$D(\omega) = H(\Theta,\omega)S(\omega), R(\omega) = H(\Phi,\omega)S(\omega)\exp(-j\omega T) \tag{2}$$

where $H(\alpha,\omega)$ is the HRTF that is the transfer function from direction α in the free field to the eardrum, $S(\omega)$ the spectrum of source sound, and T delay. Equation (2) indicates that observed spectra $D(\omega)$ and $R(\omega)$ vary with directions Θ and Φ because of the directional dependence of HRTF. Here, we assume that $D(\omega)$ and $R(\omega)$ in each pair of directions are already known by physical measurements. We find the power difference $A(f)$ from $D(\omega)$ and $R(\omega)$ using (3)

$$A(f) = 20\log_{10}\left|\frac{R(\omega)}{D(\omega)}\right| - L \tag{3}$$

where L is the level ratio in dB of the reflected sound versus direct sound used in psychological experiments. The difference between envelopes $G(f)$ is found by substituting (3) for $A(f)$ in (1). The area of comb structure P (the unit is dB $\times$ octave) is found by integrating $G(f)$ by f in a logarithmic octave scale, as (4)

$$\begin{aligned} P &= \int_{f=f_1}^{f=f_2} G(f)d(\log_2 f) \\ &= \frac{1}{\ln 2}\int_{f_1}^{f_2}\frac{G(f)}{f}\,df \end{aligned} \tag{4}$$

where range $f_1 - f_2$ Hz is the frequency range of source sound $S(\omega)$. Because our experiments are conducted as binaural tasks, total P is found by summing Ps obtained at both ears.

B. Physical Measurement

To find the area of comb structure P, we conducted physical measurements to obtain spectra $D(\omega)$ and $R(\omega)$. If we consider that only absolute $D(\omega)$ and $R(\omega)$ are required (see (3)) and delay factor $\exp(-j\omega T)$ in $R(\omega)$ [see (2)] can be omitted, we need not distinguish between direct and reflected sound in our physical measurements. If we measured spectra $D(\omega)$, it can be used as $R(\omega)$ of the same direction. All we must do is measure spectra of the four directions.

Spectra were measured at the eardrums of both ears using HATS (B&K 4128 + 4159) located at the same position as the subject of the psychological experiment. One loudspeaker

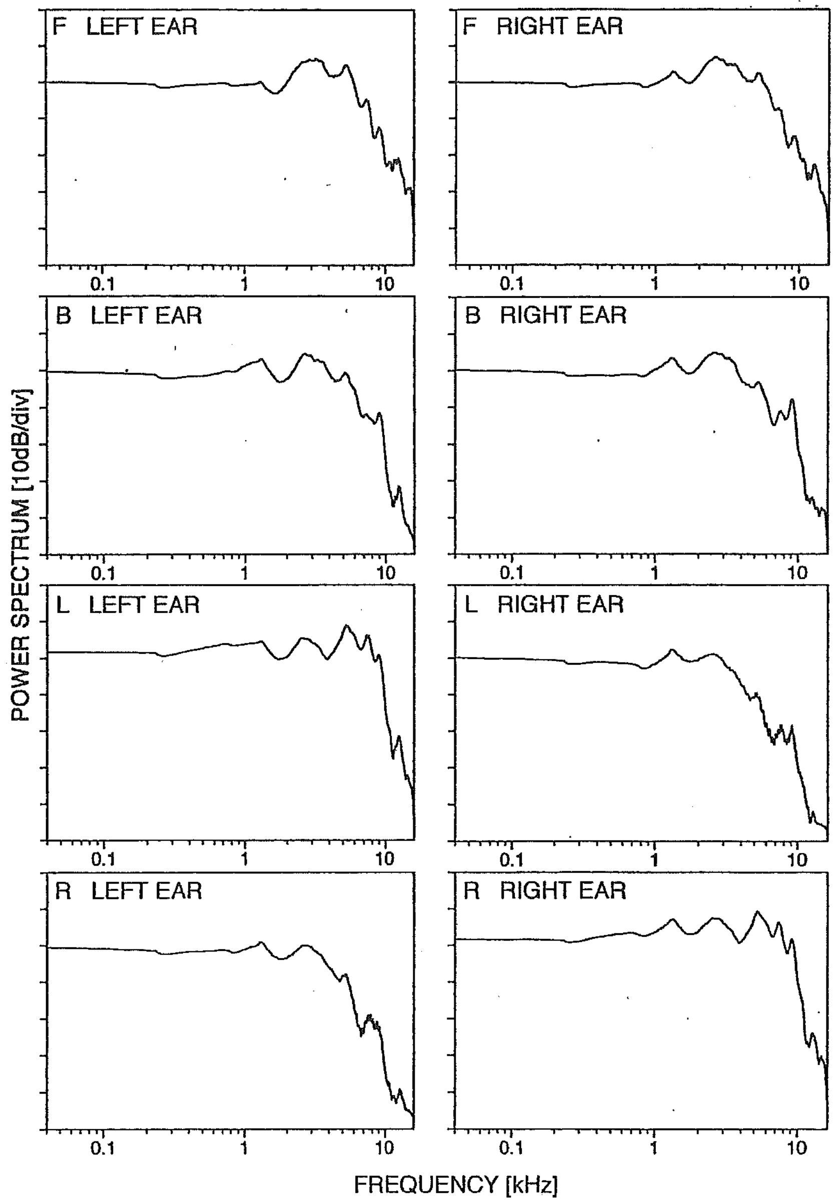


Fig. 7. Spectrum observed at the eardrum as a function of frequency. At left are four spectra at the left ear and at right four at the right ear, showing F, B, L, and R. The horizontal axis indicates the frequency (40 Hz–16 kHz) and the vertical axis the level of the power spectrum.

(BOSE 111AD) was used for directions F, B, L, and R. The probe sound was the same pink noise as used in psychological experiments. The sound pressure level was 60 dB(A) at the center of the head. Signals received by HATS were low-pass-filtered at 20 kHz 135 dB/oct cutoff by an antialias filter (NF P-86), and converted at 50 kHz sampling and 16-bit resolution by an analog-digital (AD) converter (Interface AZI-3139). Converted data was gated by a 4096-point Hamming window and FFT-analyzed by computer (NEC PC-98Ap3). One hundred results of FFT analysis were averaged to reduce noise and the resultant spectrum smoothed by a moving average with a running window 500 Hz wide on the frequency scale.

Resultant spectra for the four directions observed at eardrums of both ears (Fig. 7) tended to descend steeply as the frequency increased. One reasons for this is the property of pink noise having a -3 dB/oct weighted spectrum. Canceling the -3 dB/oct slope leaves pure HRTF. Powers in the frequency range above about 2 kHz of R at the left ear and L at the right ear are much lower than those of the other 6, because the sound is diffracted half way around the head to arrive at the ear in these two conditions and therefore loses high frequency range power through diffraction loss.

C. Calculation

From these spectra, we calculated differences between upper and lower envelopes of the comb structure $G(f)$ using (1) and (3). Fig. 8 shows examples of calculated $G(f)$ for the same conditions as in Fig. 5. To find $G(f)$ in these samples, we use the spectrum of F at the right ear is used as $D(\omega)$, and B at the right ear as $R(\omega)$.

For ratios of -12 and -22.5 dB, calculated $G(f)$ (Fig. 8) agrees well with the depth of the measured comb structure (Fig. 5). In contrast, for the ratio of 0 dB, calculated $G(f)$ appears to be much larger than that measured, especially at a

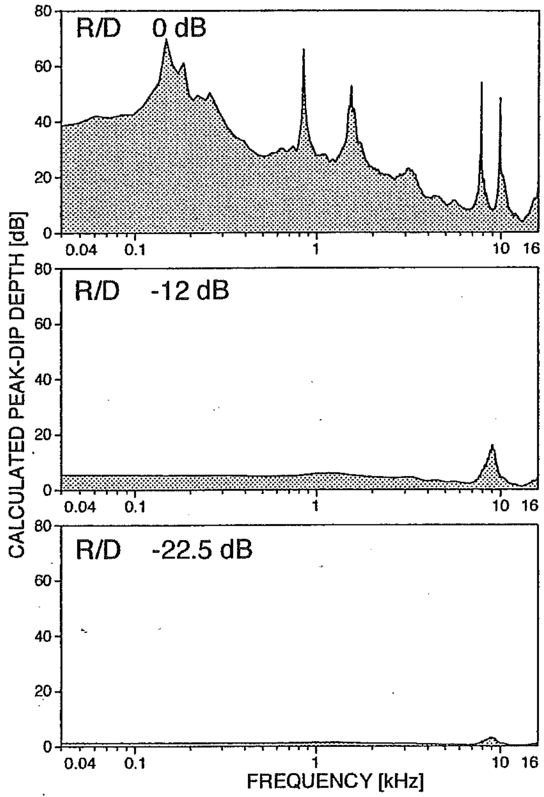


Fig. 8. Calculated peak-dip depth as a function of frequency. Above, at center, and below are the results as for the three panels of Fig. 4. The horizontal axis indicates frequency, and the vertical axis indicates the depth of the peak and dip calculated using our algorithm.

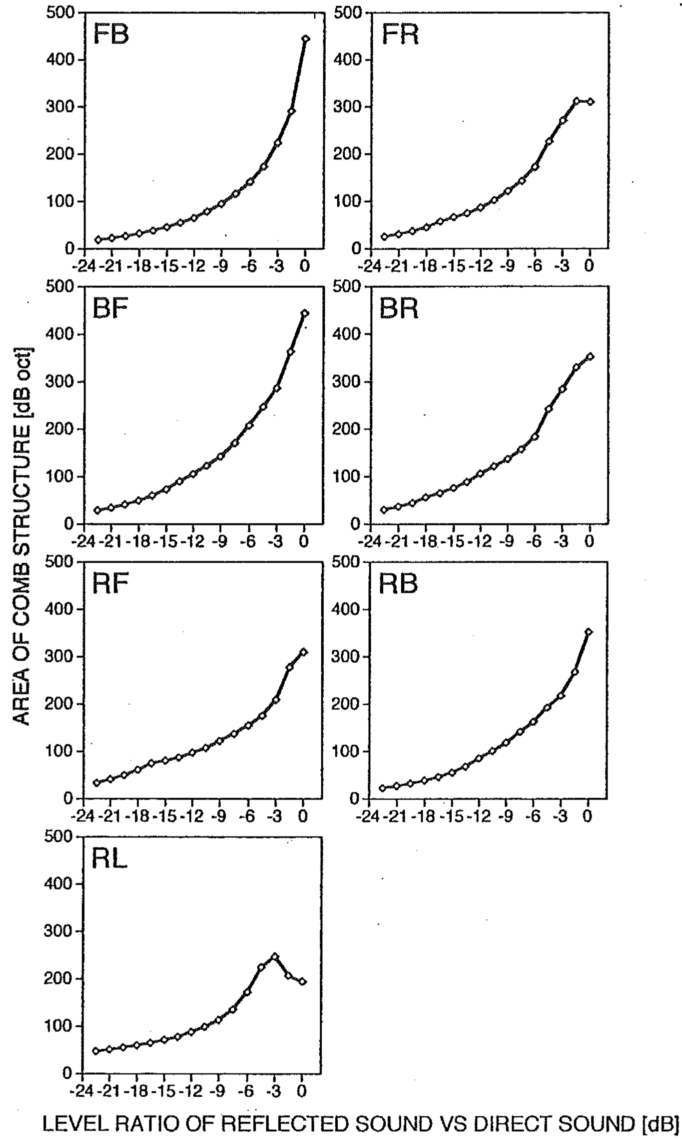


Fig. 9. Area of comb structure as a function of level ratio of reflected sound versus direct sound. The seven panels correspond to the seven pairs of directions. The areas are calculated using our algorithm.

frequency range of less than about 2 kHz. As described in III.A, the measured comb structure (Fig. 5) cannot display deep dips accurately because of the poor frequency resolution of FFT analysis. Calculated $G(f)$ (Fig. 8) represents peak-dip depth accurately and is not influenced by poor frequency resolution because our algorithm calculates depth mathematically from smoothed spectral contours and not directly from minute structures such as tips of dips influenced by poor frequency resolution.

The area of comb structure P is calculated using (4). Frequency range $f_1 - f_2$ is 40 Hz–16 kHz, yielding total P of the left and the right ears (Fig. 9).

D. Comparison With Psychological Results

We compared the calculated area of the comb structure (Fig. 9) to the perception of coloration obtained by psychological experiments (Fig. 3). Area increases monotonically as the ratio increases in all pairs except RL. For RL, the contour of the calculated area has a peak at -3 dB, as results of the perception of coloration has a peak at -4.5 dB. These comparisons indicate that at least *qualitative* tendencies of the area of the comb structure resemble those of the perception of coloration.

Now, does a *quantitative* relationship exist between these two values? Generally, a percentage in psychological detection tasks such as our psychological experiments can be represented as a psychophysical function, i.e., cumulative normal distribution, of the magnitude of physical stimulation. We plotted results for all conditions (=112 conditions = 16 ratios × 7 pairs) on normal probability paper (Fig. 10), where the vertical axis shows the perception of coloration in normal probability scale and the horizontal axis the area of the comb structure is logarithmic. The logarithm of the area of the comb structure and z (variable of cumulative standard normal distribution) of the perception of coloration are quite linearly related, with a correlation coefficient of 0.949 (Fig. 10), enabling the relation to be represented approximately as in (5)

$$Q = F(a \ln P + b) \tag{5}$$

where Q% is the perception of coloration, $F(z)$% cumulative standard normal distribution, and a, b constants. For our experiments, constants are found to be $a = 0.919$ and $b = -4.341$ using regression analysis.

In perceptions of coloration predicted using (5), (Fig. 11), predicted values agree well with results from psychological experiments.

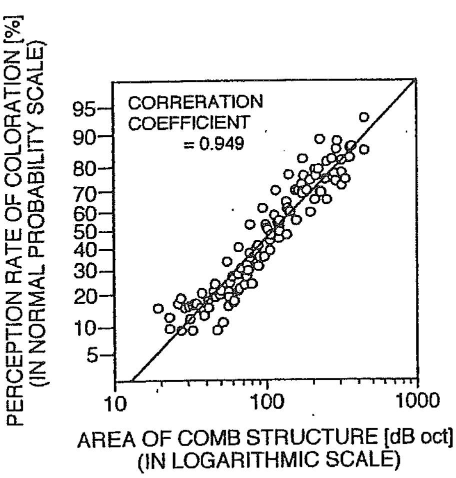


Fig. 10. Relationship between perception of coloration and the area of comb structure. The horizontal axis indicates the area calculated using our algorithm logarithmically and the vertical axis shows perception obtained by psychological experiments on a normal probability scale. Results of the 112 conditions (=16 ratios × 7 pairs) are plotted. Note the linear relationship between area and perception. The correlation coefficient is 0.949.

IV. Discussion

According to the results of our psychological experiments (Fig. 3), the 50% threshold of the perception of coloration is −10 dB for all seven pairs and does not appear to depend on the direction of sounds. From the viewpoint of "threshold," we concluded that the perception of coloration does not depend on direction, as Boussard *et al.* [15] reported. Ando *et al.* [1] reported that threshold improves slightly as reflected sound direction moves from a 0° to a 30° azimuth, but their experiments did not cancel out spatial cues such as changes in sound images, so the improvement of the threshold does not appear to be caused by spectral cues alone.

In contrast, the dependence on direction appears as the ratio approaches 0 dB, with RL significantly different from the other six pairs because both ears should observe diffracted sound in RL, whereas at least one of the two ears observes nondiffracted sound in the other six pairs. Our results suggest that, if direct sound comes from a lateral direction and reflected sound from the opposite lateral direction, the perception of coloration does not increase monotonically even if the ratio approaches 0 dB.

We next proposed a numerical model explaining results of the perception of coloration depending on direction and the ratio using objective spectrum analysis. We defined the area of the comb structure on the spectrum, proposing an algorithm to calculate it mathematically. Comparing the calculated area to the perception of coloration, we found that the logarithm of the area and z of perception are linearly related. We proposed a model based on the relationship, demonstrating that our model predicts the perception of coloration well, especially the nonmonotonic increase of RL and monotony of the others.

We conducted psychological experiments in binaural listening in which interaural information such as interaural time difference (ITD), interaural level difference (ILD), and interaural correlation coefficient can be influenced by a subject's performance. Our model can predict results of experiments without considering these, however, because subjects paid attention only to the change in timbre in psychological experiments, and not to the change in the sound image based on interaural information. The results of our psychological experiments therefore do not include the influence of the interaural information and does not take them into account except for totaling areas at both ears.

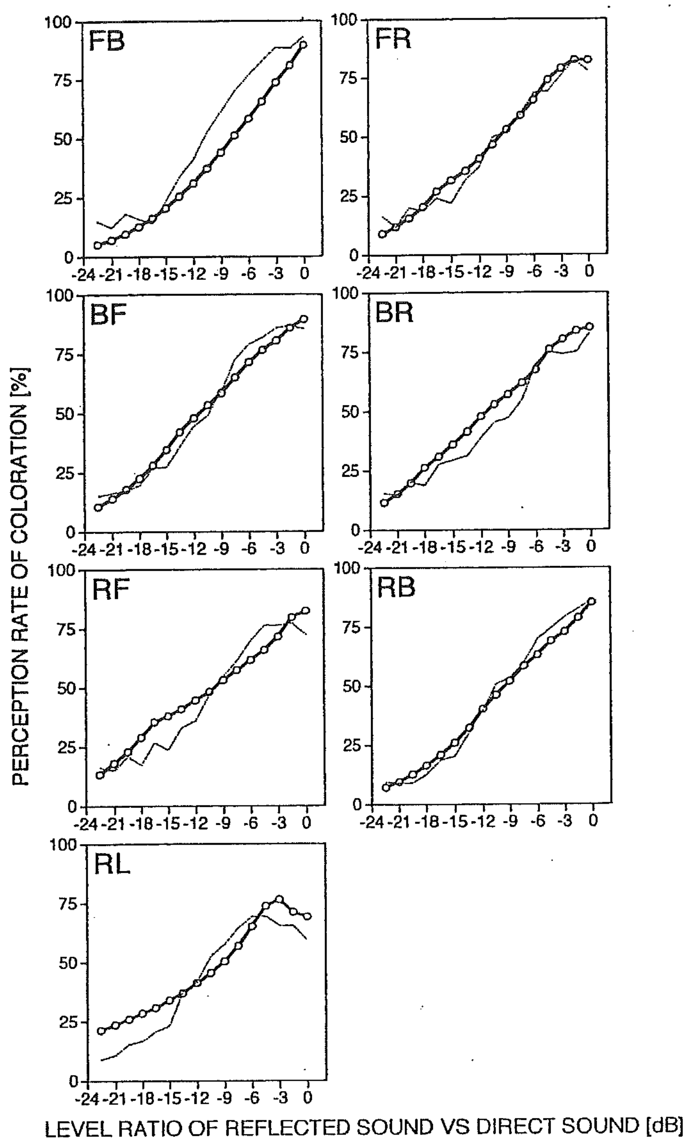


Fig. 11. Predicted perception of coloration as a function of the level ratio of reflected versus direct sound. The seven panels shown correspond to the seven pairs of directions. The solid line shows predicted values and the dotted line results from psychological experiments (same as Fig. 3).

Our model is similar to the method of Meynial *et al.* [14] in that it focuses on spectral distribution. They used stochastic deviation as the index of spectral distribution. Because their method is to predict coloration in a reverberation room with many comb structures layered on the spectrum, spectral distribution must be treated as stochastic. In contrast, our model predicts coloration in a simple sound field with only one reflected sound and the comb structure observed clearly, making our model is simpler than theirs.

We expect prediction will become more accurate if we take into account the following not yet considered in our model.

1) To simulate auditory scale, the Bark scale is better for the frequency scale rather than the logarithmic.
2) Spectra of sound should be measured at every subject's ears, because HRTF differs individually among subjects, and HRTF of HATS is not exactly the same as that of subjects.
3) To measure perception of *pure* coloration caused by the comb structure on the spectrum, we should cancel binaural pitch effects such as Huggins pitch (cf. [17]), and cancel the precedence effect in the psychological experiment.

Given that the perception of coloration becomes stronger as delay T gets shorter (e.g., [1]), constants a and b in (5) must vary depending on T.

V. Conclusion

We studied the relationship between the perception of coloration and the direction of direct and reflected sounds, proposing a numerical model to explain results of the perception of coloration depending on direction and the ratio of the reflected sound versus direct sound using objective spectrum analysis, with the following results.

1) The 50% threshold of coloration does not appear to differ depending on sound direction, agreeing with the previous study.
2) As the ratio approaches 0 dB, a difference appears depending on direction. If direct sound comes from a lateral direction and reflected sound from the opposite lateral direction, the perception of coloration does not increase monotonically even if the ratio approaches 0 dB, whereas the other pairs of the directions show an approximately monotonic increase.
3) We assumed that the difference depending on direction resulted from the directional dependence of the spectrum including HRTF, proposing a numerical model for predicting perceptions of coloration using the area of the comb structure on the spectrum observed at the eardrum. We demonstrated that our model predicts them well.

References

[1] Y. Ando and H. Alrutz, "Coloration perception in sound fields in relation to the autocorrelation function," *J. Acoust. Soc. Amer.*, vol. 71, pp. 616–618, Mar. 1982.

[2] F. A. Bilsen and R. J. Ritsma, "Some parameters influencing the perceptibility of pitch," *J. Acoust. Soc. Amer.*, vol. 47, pp. 469–475, 1970.

[3] J. M. Kates, "A central spectrum model for the coloration perception in filtered Gaussian noise," *J. Acoust. Soc. Amer.*, vol. 77, pp. 1529–1534, Apr. 1985.

[4] I. G. Bassett and E. J. Eastmond, "Echolocation: Measurement of pitch versus distance for sounds reflected from a flat surface," *J. Acoust. Soc. Amer.*, vol. 36, pp. 911–916, May 1964.

[5] W. R. Thurlow and A. M. Small Jr., "Pitch perception for certain periodic auditory stimuli," *J. Acoust. Soc. Amer.*, vol. 27, pp. 132–137, Jan. 1955.

[6] W. R. Thurlow, "Further observation on pitch associated with a time difference between 2 pulse trains," *J. Acoust. Soc. Amer.*, vol. 29, pp. 1310–1311, Dec. 1957.

[7] W. A. Yost and R. Hill, "Strength of the pitch associated with ripple noise," *J. Acoust. Soc. Amer.*, vol. 64, pp. 485–492, Aug. 1978.

[8] ——, "Models of the pitch and pitch strength of ripple noise," *J. Acoust. Soc. Amer.*, vol. 66, pp. 400–410, Aug. 1979.

[9] W. A. Yost, "The dominance region and ripple noise pitch: A test of the peripheral weighting model," *J. Acoust. Soc. Amer.*, vol. 72, pp. 416–425, Aug. 1982.

[10] R. M. Warren, J. A. Bashford Jr., and J. M. Wrightson, "Infrapitch echo," *J. Acoust. Soc. Amer.*, vol. 68, pp. 1301–1305, Nov. 1980.

[11] W. A. Yost and M. J. Moore, "Temporal changes in a complex spectral profile," *J. Acoust. Soc. Amer.*, vol. 81, pp. 1896–1905, June 1987.

[12] R. M. Warren and J. A. Bashford Jr., "Broadband repetition pitch: Spectral dominance or pitch averaging?," *J. Acoust. Soc. Amer.*, vol. 84, pp. 2058–2062, Dec. 1988.

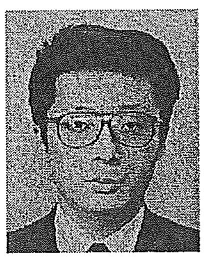
[13] J. R. Pierce, "Periodicity and pitch perception," *J. Acoust. Soc. Amer.*, vol. 90, pp. 1889–1893, Oct. 1991.

[14] X. Meynial and O. Vuichard, "Objective measure of sound coloration in rooms," *Acoustica*, vol. 85, pp. 101–107, 1999.

[15] P. Boussard and F. Santon, "Perception of coloration of a white noise by a lateral echo," *J. Phys. IV*, vol. 2, pp. 221–224, Apr. 1992. Colloque C1.

[16] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition.* Cambridge, MA: MIT Press, 1997, pp. 201–287.

[17] B. C. J. Moore, *Hearing*, 2nd ed. New York: Academic, 1995, pp. 267–295.



Yoshikazu Seki was born in 1966. He received the Ph.D. degree in engineering from Hokkaido University, Japan, in 1994.

He is currently the Senior Research Scientist of Institute for Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology (AIST), Japan. His research interests include training method and assistive technology for the blind people by using auditory obstacle perception, sound localization, echo perception, or other auditory cues.

Dr. Seki is a member of the Acoustical Society of Japan, Japanese Association of Rehabilitation for the Visually Impaired, Rehabilitation Engineering Society of Japan, and Association of Sensory Substitution, Japan. He is also Secretary of Association of Sensory Substitution, Japan, and a part-time instructor of the College of National Rehabilitation Center for the Persons with Disabilities, Japan.

Kiyohide Ito was born in Japan with congenital visual impairment in 1964. He received the Ph.D. degree from University of Tsukuba, Japan, in 1998.

He is currently Assistant Professor at Future University, Hakodate, Japan. He majors in cognitive psychology and ecological psychology. He is interested in auditory obstacle perception, echolocation, and auditory information for aid to blind persons. He has been engaged in studies to demonstrate what kind of auditory information blind persons use to control their actions and how they perceptually learn to use reflected sound for detection of environment.

RESNA 27TH INTERNATIONAL ANNUAL CONFENCE

Technology & Disability: Research, Design, Practice & Policy

June 18 to June 22, 2004
Orlando, Florida

OBSTACLE PERCEPTION TRAINING SYSTEM AND CD FOR THE BLIND

Yoshikazu Seki and Kiyohide Ito
National Institute of Advanced Industrial Science and Technology
Tsukuba , Ibaraki , 305-8566 Japan

ABSTRACT

Obstacle perception is a skill to detect presence of "silent" object, such as wall, pole, etc., by perceiving the acoustical cues, such as reflected sound, etc., through auditory sense. This skill is very important for orientation and mobility (O&M) of the blind. We are studying the training system for acquiring this skill in the blind education and rehabilitation by using acoustical technologies. Our training system consists of sound processors for making reflected sounds and several loudspeakers, and can reproduce ideal sound fields for learning the principle of obstacle perception. We are also distributing the audio CD that contains these sound fields to the people concerned with the blind education and rehabilitation. Our system is now used in the school for O&M instructor in Japan , and our CDs have been distributed to about 15 0 Japanese facilities concerned with the blind.

KEYWORDS

B lind, O&M, obstacle perception, echolocation, auditory training

BACKGROUND

Obstacle perception (or obstacle sense, human echolocation) is an experiential ability to detect presence of "silent" object, such as wall, pole, etc., which does not make any sounds, by perceiving the acoustical cues, such as reflection or diffraction of environmental or self-generated sounds, through auditory sense. This skill is very important for orientation and mobility (O&M) of the blind (1) . Learning the relation between the existence of object and the variation of sound that is caused by the object enables to master the obstacle sense.

Conventional training for acquiring obstacle perception in the blind education or rehabilitation is usually performed in real environment. The O&M instructor instructs the blind trainee to approach the wall, or bring a small board near to the trainee's face, to show the sound variation that is caused by presence of object. The trainee should learn the cue of obstacle perception by hearing the shown sound variation experientially.

RESEARCH QUESTION

However, in the real environment, the sound variation caused by the object can hardly be heard, because of irregular variation of environmental sounds, attenuation of the reflected sounds, or being disturbed by other sounds. The conventional training in the real environment is sometimes difficult for beginners of obstacle perception. In order to provide the ideal training environment for the beginners, the new acoustical technologies that can reproduce an ideal sound field variation artificially should be introduced into the blind education or rehabilitation.

In this paper, we report our new training system or training CD for acquiring obstacle perception by using acoustical technologies .

METHOD

As the first step of applying the acoustical technology to the training of obstacle perception, we developed a new acoustical training system in 1998 (2) . Our system consists of the signal processors and the loudspeakers that are arranged in a ring, and can reproduce the sound field variation by emitting the delayed sounds with respect to the direct sounds, which simulates reflected sounds from wall. Our system produces the ideal sound field variation by assuming that (i) the reflected sounds are perfectly (not attenuated) reflected at the surface of the object, (ii) the diffracted sounds are completely attenuated, and (iii) no other sound exists in the sound field.

The first model of our system had 8 digital delays (Roland SDE-330), and 16 loudspeakers (BOSE 111AD). The digital delays are controlled by the computer (NEC PC-9821Ap3), and make the reflected sounds by delaying the original sound signal. The delay times are calculated by the computer, and informed to the digital delays through MIDI . The original sound signal that simulates the environmental sound is white noise, pink noise, or true-recorded sound. The sound signal was recorded on multi track recorders (YAMAHA MT4X).

The first system can emit up to 8 reflected sounds, and then project "virtual wall." The virtual wall can change distance, direction, and width. The distance is adjusted by the delay times. The direction and width are adjusted by assignment of the reflected sound loudspeakers.

Many O&M instructors and the people concerned with blind education or rehabilitation have tried this system, and they experienced that our system can reproduce the ideal sound variation that is easier to understand for beginners than conventional training.

However, our first system was too expensive and too large to be introduced into the facilities for blind education or rehabilitation. So we also developed the simplified version that consists of 1 or 2 digital delays and 2 or 4 loudspeakers. The simplified version is now working in the Coll e ge of National Rehabilitation Center for the Persons with Disab ilities , Japan for educational use.

Our training system, even if it is simplified, requires economic load and space to set its hardware, when it is introduced into the blind facilities. In order to reduce economic load and save space of the blind facilities, we developed the training audio CD that contains ideal sound field data in 2001 - 2002, and are now distributing it free of charge.

Our training CD can be reproduced by general audio equipment for home use, and it can also generate the ideal sound field variation as well as the system. Our training CD contains one reflected sound in the right channel and one direct sound in the left channel. It is just the same as the simplified system of 1 delay and 2 loudspeakers.

When play this CD, the two stereophonic loudspeakers are arranged 2.4 - 3.0 m apart and facing each other. The listener's head is at the center of the two loudspeakers. The "virtual wall" appears in the direction of the right channel loudspeaker.

The sound field data were calculated by sound processing software (developed by REAL Software REALbasic) in computer (Apple PowerMac G4), and are recorded by CD burner (Apple iTunes & SuperDrive).

We first developed the trial version of the training CD, "Version 0.0" in 2001 and distributed it to over 50 facilities concerned with blind education or rehabilitation in Japan . Then we modified all data by their opinions, and developed the next version "Version 1.0" in 2002.

"Version 1.0" is focused on the usefulness in practical training scene, and contains various situations. The original sound signal that simulates the environmental sound is white noise, or true-recorded sound. The sound data simulate the wall movement in various ways. For example, wall approaches, approaches and goes away, jumps, and moves randomly. Total of 21 movement patterns are simulated.

CD of "Version 1.0" is a single size (8 cm) audio CD, and contains 42 tracks data. Total recording time is 21 minutes 49 seconds. We are now distribution it to the facilities of blind education or rehabilitation free of charge. We are also providing the free download site of the sound field data in WAVE file format on internet. The information about the training CD, the WAVE files, and the training manual are available on our web site : http://staff.aist.go.jp/yoshikazu-seki/ CD/CD10/

RESULTS

In this paper, we reported our training system or training CD for acquiring obstacle perception. Many O&M instructors and the people concerned with blind education or rehabilitation have tried our training system, and they reported that our system can reproduce the ideal sound variation that is easier to understand for beginners than conventional training. The simplified version of our training system is now working in the Coll e ge of National Rehabilitation Center for the Persons with Disab ilities , Japan for educational use. The latest version of our training CD "Version 1.0" has been distributed to over 15 0 blind facilities.

REFERENCES

1. Blasch BB , Wiener WR , and Welsh RL , Editors, " Foundations of Orientation and Mobility Second Edition, " AFB Press, New York , 1997.
2. Seki Y , "Systematic Auditory Training of Obstacle Sense for the Visually Impaired by using Acoustical VR System," Human-Computer Interaction 2, 999-1003 (1999) .

Author Contact Information:

Yoshikazu Seki , Ph.D. ,
Perception Group,
Institute for Human Science and Biomedical Engineering,
National Institute of Advanced Industrial Science and Technology ,
1-1 Higashi, Tsukuba, Ibaraki 305-8566 J apan ,
Office Phone: +81-29-861-6716,

E MAIL : yoshikazu-seki@aist.go.jp ,
URL : http://staff.aist.go.jp/yoshikazu-seki/

AUDITORY OBSTACLE PERCEPTION TRAINING SYSTEM AND CD FOR THE BLIND

Y. Seki, & K. Ito***
**National Institute of Advanced Industrial Science and Technology,*
1-1 Higashi, Tsukuba, Ibaraki 305-8566 Japan.
Tel. +81 29 861 6716. Fax: +81 29 861 6761. Email: yoshikazu-seki@aist.go.jp
***Future University-Hakodate, Japan.*
Email: itokiyo@fun.ac.jp

Abstract:

Obstacle perception is a skill to detect presence of "silent" object, such as wall, pole, etc., by perceiving the acoustical cues, such as reflected sound, etc., through auditory sense. This skill is very important for orientation and mobility (O&M) of the blind. We are studying the training system for acquiring this skill in the blind education and rehabilitation by using acoustical technologies. Our training system consists of sound processors for making reflected sounds and several loudspeakers, and can reproduce ideal sound fields for learning the principle of obstacle perception. We are also distributing the audio CD that contains these sound fields to the people concerned with the blind education and rehabilitation. Our system is now used in the school for O&M instructor in Japan, and our CDs have been distributed to about 150 Japanese facilities concerned with the blind.

Keywords:

blind, O&M, obstacle perception, echolocation, auditory training

Table of Contents

1. Introduction

Obstacle perception (or obstacle sense, human echolocation) is an experiential ability to detect presence of "silent" object, such as wall, pole, etc., which does not make any sounds, by perceiving the acoustical cues, such as reflection or diffraction of environmental or self-generated sounds, through auditory sense (Figure 1). This skill is very important for orientation and mobility (O&M) of the blind (Blasch, et al., 1997). Learning the relation between the existence of object and the variation of sound that is caused by the object enables to master the obstacle sense.

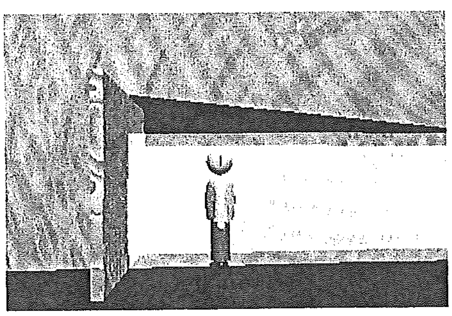


Figure 1, Schematic Explanation of "Obstacle Perception."
Blind person can detect presence of object, such as wall, by hearing reflection or diffraction of environmental or self-generated sounds through auditory sense. This ability is acquired by learning.

In this paper, we report our new training system or training CD for acquiring obstacle perception by using acoustical technologies.

* Back to Table of Contents.

2. Previous Training

Conventional training for acquiring obstacle perception in the blind education or rehabilitation is usually performed in real environment. The O&M instructor instructs the blind trainee to approach the wall (Left panel of Figure 2), or brings a small board near to the trainee's face (Right panel of Figure 2), to show the sound variation that is caused by presence of object. The trainee should learn the cue of obstacle perception by hearing the shown sound variation experientially.

However, in the real environment, the sound variation caused by the object can hardly be heard, because of irregular variation of environmental sounds, attenuation of the reflected sounds, or being disturbed by other sounds. The conventional training in the real environment is sometimes difficult for beginners of obstacle perception. In order to provide the ideal training environment for the beginners, the new acoustical technologies that can reproduce an ideal sound field variation artificially should be introduced into the blind education or rehabilitation.

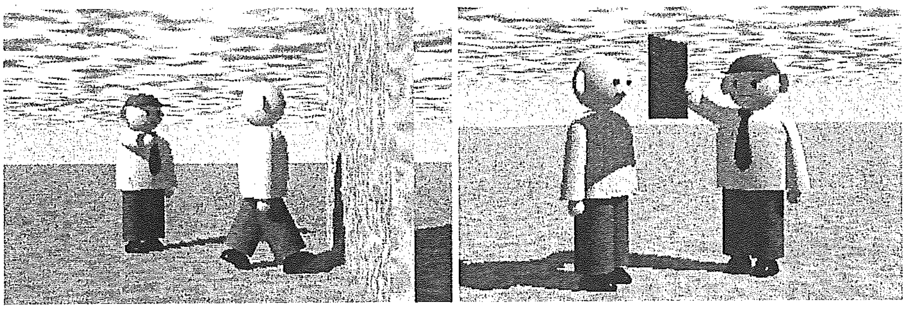


Figure 2, Schematic Explanation of Conventional Training of Obstacle Perception.
The blind trainee should learn the cue of obstacle perception by hearing the sound variation experientially with approaching wall or board.

* Back to Table of Contents.

3. Training System

As the first step of applying the acoustical technology to the training of obstacle perception, we developed a new acoustical training system (Seki, 1998). Our system consists of the signal processors and the loudspeakers that are arranged in a ring, and can reproduce the sound field variation by emitting the delayed sounds with respect to the direct sounds, which simulates reflected sounds from wall. Our system produces the ideal sound field variation by assuming that (i) the reflected sounds are perfectly (not attenuated) reflected at the surface of the object, (ii) the diffracted sounds are completely attenuated, and (iii) no other sound exists in the sound field.

The first model of our system had 8 digital delays (Roland SDE-330), and 16 loudspeakers (BOSE 111AD). The digital delays are controlled by the computer (NEC PC-9821Ap3), and make the reflected sounds by delaying the original sound signal. The delay time of the reflected sound ΔT is depending on both distance d and sound angle of incidence to the wall ϑ, and is found by Equation (1):

$\Delta T = 2\ d \cos \vartheta\ /\ c$ --- (1)

where c is the sound velocity in the air (= 340 m/s). The delay times are calculated by the computer, and informed to the digital delays through MIDI. The original sound signal that simulates the environmental sound is white noise, pink noise, or true-recorded sound. The sound signal was recorded on multi track recorders (YAMAHA MT4X). (Figure 3, and 4)

The first system can emit up to 8 reflected sounds, and then project "virtual wall." The virtual wall can change distance, direction, and width. The distance is adjusted by the delay times. The direction and width are adjusted by assignment of the reflected sound loudspeakers.

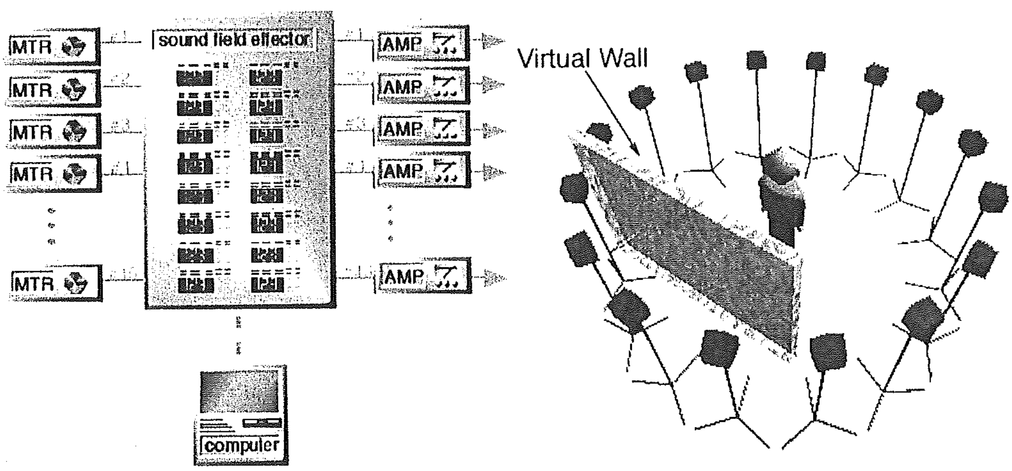


Figure 3, Outlook of Our Acoustical Training System for Obstacle Perception.
This system consists of signal processor (left panel) and loudspeakers (right panel). This system can reproduce ideal sound field variation for he beginners' training. The wall shown in right panel is "virtual wall" projected by our system.

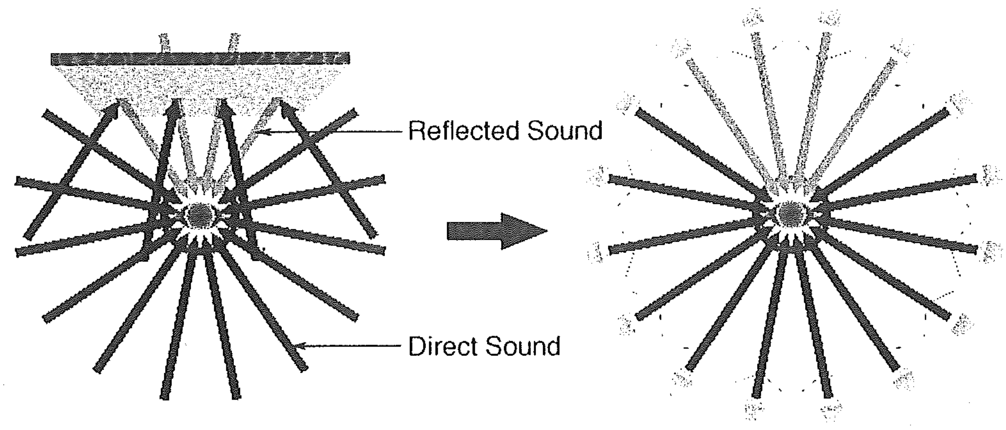


Figure 4, Principle of reproduction of sound field variation. Our system simulates reflected sounds that reflect at surface of wall. Delay time of each reflected sound is calculated by computer, and processed by digital delays. Left panel shows environmental sound field where some sounds reflect at wall, and right panel shows simulated sound field by our loudspeaker system.

Many O&M instructors and the people concerned with blind education or rehabilitation have tried this system, and they experienced that our system can reproduce the ideal sound variation that is easier to understand for beginners than conventional training.

However, our first system was too expensive and too large to be introduced into the facilities for blind education or rehabilitation. So we also developed the simplified version that consists of 1 or 2 digital delays and 2 or 4 loudspeakers. The simplified version is now working in the College of National Rehabilitation Center for the Persons with Disabilities, Japan for educational use.

* Back to Table of Contents.

4. Training CD

Our training system, even if it is simplified, requires economic load and space to set its hardware, when it is introduced into the blind facilities. In order to reduce economic load and save space of the blind facilities, we developed the training audio CD that contains ideal sound field data, and are now distributing it free of charge.

Our training CD can be reproduced by general audio equipment for home use, and it can also generate the ideal sound field variation as well as the system. Our training CD contains one reflected sound in the right channel and one direct sound in the left channel. It is just the same as the simplified system of 1 delay and 2 loudspeakers.

When play this CD, the two stereophonic loudspeakers are arranged 2.4 - 3.0 m apart and facing each other. The listener's head is at the center of the two loudspeakers. The "virtual wall" appears in the direction of the right channel loudspeaker. (Right panel of Figure 5)

The sound field data were calculated by sound processing software (developed by REAL Software REALbasic) in computer (Apple PowerMac G4), and are recorded by CD burner (Apple iTunes & SuperDrive).

We first developed the trial version of the training CD, "Version 0.0" in 2001 and distributed it to over 50 facilities concerned with blind education or rehabilitation in Japan. Then we modified all data by their opinions, and developed the next version "Version 1.0" in 2002. (Left panel of Figure 5)

"Version 1.0" is focused on the usefulness in practical training scene, and contains various situations. The original sound signal that simulates the environmental sound is white noise, or true-recorded sound. The sound data simulate the wall movement in various ways. For example, wall approaches, approaches and goes away, jumps, and moves randomly. Total of 21 movement patterns are simulated. (Figure 6, Table 1)

CD of "Version 1.0" is a single size (8 cm) audio CD, and contains 42 tracks data. Total recording time is 21 minutes 49 seconds. We are now distributing it to the facilities of blind education or rehabilitation free of charge. We are also providing the free download site of the sound field data in WAVE file format on internet. The information about the training CD, the WAVE files, and the training manual are available on our web site: http://staff.aist.go.jp/yoshikazu-seki/CD/CD10/ (Seki, 2002).

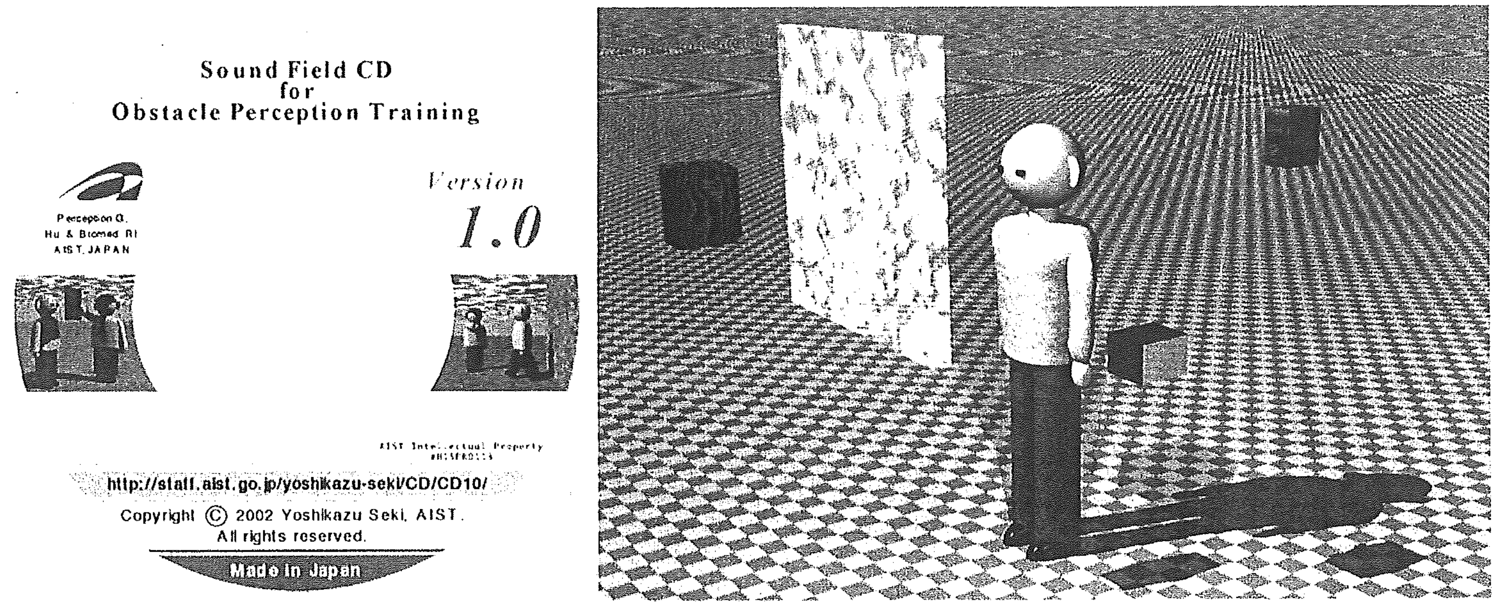


Figure 5, Outlook of Our Training CD "Sound Field CD for Obstacle Perception Training, Version 1.0," and Schematic Explanation of Reproduction of Sound Field.
This CD can be reproduced by home use audio equipment. Two stereophonic loudspeakers are arranged 2.4 - 3.0 m apart and facing each other. Listener's head is at center of two loudspeakers. "Virtual wall" appears in direction of right channel loudspeaker.

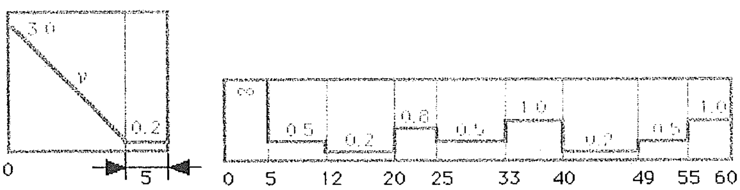


Figure 6, Examples of the Simulated Wall Movement in the Training CD "Version 1.0." Horizontal axis is time in second, and vertical axis is distance in meter. Left panel shows the approaching movement of the wall ("v" means constant velocity), and right panel shows one example of the random movements. Total of 21 movement patterns are simulated in "Version 1.0."

The following list shows the contents of "Version 1.0."

- 01 : White : Move from 3m to 0.2m in 0.7m/s : 9s
- 02 : White : Move from 3m to 0.2m in 0.4m/s : 12s
- 03 : White : Move from 3m to 0.2m in 0.2m/s : 19s
- 04 : White : Move between 3m and 0.2m in 0.7m/s : 56s
- 05 : White : Move between 3m and 0.2m in 0.4m/s : 56s
- 06 : White : Move between 3m and 0.2m in 0.2m/s : 56s
- 07 : White : Jump from infinity to 2.0m : 15s
- 08 : White : Jump from infinity to 1.5m : 15s
- 09 : White : Jump from infinity to 1.0m : 15s
- 10 : White : Jump from infinity to 0.9m : 15s
- 11 : White : Jump from infinity to 0.8m : 15s
- 12 : White : Jump from infinity to 0.7m : 15s
- 13 : White : Jump from infinity to 0.6m : 15s
- 14 : White : Jump from infinity to 0.5m : 15s
- 15 : White : Jump from infinity to 0.4m : 15s
- 16 : White : Jump from infinity to 0.3m : 15s
- 17 : White : Jump from infinity to 0.2m : 15s
- 18 : White : Random (Slant wall) : 60s
- 19 : White : Random (Rough wall) : 60s
- 20 : White : Random (Exits) : 60s
- 21 : White : Random (Poles) : 60s
- 22 : Environmental : Move from 3m to 0.2m in 0.7m/s : 9s
- 23 : Environmental : Move from 3m to 0.2m in 0.4m/s : 12s
- 24 : Environmental : Move from 3m to 0.2m in 0.2m/s : 19s
- 25 : Environmental : Move between 3m and 0.2m in 0.7m/s : 56s
- 26 : Environmental : Move between 3m and 0.2m in 0.4m/s : 56s
- 27 : Environmental : Move between 3m and 0.2m in 0.2m/s : 56s
- 28 : Environmental : Jump from infinity to 2.0m : 15s
- 29 : Environmental : Jump from infinity to 1.5m : 15s
- 30 : Environmental : Jump from infinity to 1.0m : 15s
- 31 : Environmental : Jump from infinity to 0.9m : 15s
- 32 : Environmental : Jump from infinity to 0.8m : 15s
- 33 : Environmental : Jump from infinity to 0.7m : 15s
- 34 : Environmental : Jump from infinity to 0.6m : 15s
- 35 : Environmental : Jump from infinity to 0.5m : 15s
- 36 : Environmental : Jump from infinity to 0.4m : 15s
- 37 : Environmental : Jump from infinity to 0.3m : 15s