応用論文

TVRSJ Vol.3 No.1, 1998

マルチスピーカ方式による2音源立体音像制御

天野 克巳、[*1] 松下 文雄、[*2] 柳川 博文、[*3]
公園 マイケル、[*4] イエンツ ヘルダー、[*5] ウイリアム L. マーテンス、[*6]
木場 美治、東山 三樹夫[*7]

A Virtual Reality Sound System Using Room-Related Transfer Functions Delivered Through a Multispeaker Array: The PSFC at the University of Aizu Multimedia Center

Katsumi Amano,[*1] Fumio Matsushita,[*2] Hirofumi Yanagawa,[*3]
Michael Cohen,[*4] Jens Herder,[*5] William L. Martens,[*6]
Yoshiharu Koba and Mikio Tohyama[*7]

Abstract — PSFC は会津大学マルチメディアセンターに設置された，DSP と半球面上に配置されたスピーカ群からなる音場制御システムである．本システムは2つの音源をリアルタイムに仮想空間の任意の位置に音像定位させ，同時に仮想空間の残響時間，大きさ，ライブネスを制御できる．交互モードでは仮想音源の方向のみ制御する．すなわち，間接音を生成しないので，音源までの距離感はラウドネスによる以外変化しない．スピーカ群が構成する半球面の直径はおよそ 10 m である．これにより約２０人が，微妙な音場感を聴取できる最適聴取位置近傍にすわるか立つことができ，最多で約５０人が視聴できる．リアプロジェクション型の立体画像システムと組み合わせることにより，本システムはより進歩したマルチメディアおよび仮想現実への応用を可能にしている．

The PSFC is a DSP-driven hemispherical loudspeaker array, installed at the University of Aizu Multimedia Center. It features realtime manipulation of the primary components of sound spatialization for each of two audio sources located in a virtual environment, including both the content (source location: apparent direction and distance) and context (room characteristics: room size and liveness). In an alternate mode, it can also direct the destination of the two separate input signals across 14 loudspeakers, manipulating the apparent direction of the virtual sound sources with no control over apparent distance other than that afforded by source loudness (i.e., no simulated environmental reflections or reverberation). The PSFC speaker dome is about 10 m in diameter, accommodating about fifty simultaneous users, including about twenty users comfortably standing or sitting near its "sweet spot," the area in which the illusions of sound spatialization are most vivid. Collocated with a large screen rear-projection stereographic display, the PSFC is intended for advanced multimedia and *v*irtual *r*eality applications.

Keywords : audio signal processing, audio telecommunication, auralization, calm technology, directional mixing console, multichannel sound reproduction, room-related transfer functions, roomware, sound localization, virtual conferencing environment

*1:〒 350-02; 埼玉県 鶴ヶ島市 富士見 6-1-1; パイオニア (株) 総合研究所; 音響システム研究部 ・第１研究室; ama@crdl.pioneer.co.jp
*2:〒 350; 埼玉県川越市山田字西町 25-1; パイオニア株式会社; 川越工場; 第３商品開発部 開発３課; mat@3s.kwg.pioneer.co.jp
*3:〒 275; 千葉県習志野市津田沼 2-17-1; 千葉工業大学; 情報ネットワーク学科; yanagawa@net.it-chiba.ac.jp
*4:〒 965-8580; 会津大学; ヒューマン インタフェース 講座; スペーシャル メディア グループ; mcohen@u-aizu.ac.jp
*5:〒 965-8580; 会津大学; コンピュータ 産業学 講座; スペーシャル メディア グループ; herder@u-aizu.ac.jp
*6:〒 965-8580; 会津大学; ヒューマン インタフェース 講座; スペーシャル メディア グループ; wlm@u-aizu.ac.jp
*7:〒 192; 東京都八王子市中野町２６６５－１; 工学院大学; 八王子キャンパス ５－６０１; ctdeneuv@sin.cc.kogakuin.ac.jp
*1:Acoustic Systems Dept./Lab. 1; Corporate Research & Development Lab; Pioneer Electronic Corp.; 6-1-1 Fujimi Tsurugashima-shi; Saitama 350-02; ama@crdl.pioneer.co.jp
*2:Dev. Section 3; Product Development Dept. 3; Mobile Entertainment Comp.; Pioneer Electronic Corp.; 25-1 Aza-Nishimachi Yamada Kawagoe-shi; Saitama 350; mat@3s.kwg.pioneer.co.jp
*3:Network Acoustics Laboratory; Dept. of Network Science; Chiba Inst. of Technology; 2-17-1 Tsudanuma; Narashino-shi; Chiba 275; yanagawa@net.it-chiba.ac.jp
*4:Spatial Media Group; Human Interface Lab; Univ. of Aizu 965-8580; mcohen@u-aizu.ac.jp
*5:Spatial Media Group; Computer Industry Lab; Univ. of Aizu 965-8580; herder@u-aizu.ac.jp
*6:Spatial Media Group; Human Interface Lab; Univ. of Aizu

1. Introduction

As suggested by Table 1, a "spatially immersive" system integrates the embedded experience of immersive systems with the convenience of so-called "fishtank" or "desktop VR" systems, allowing multiple co-located users to share perspective [14] [9].

Fig. 1 is a schematic of the PSFC (for **P**ioneer **S**ound **F**ield **C**ontroller) [1], a DSP-driven hemispherical loudspeaker array installed in the Synthetic World Zone at the University of Aizu Multimedia Center[1]. A **r**oom-**r**elated **t**ransfer **f**unction, in analogy to an HRTF (**h**ead-**r**elated **t**ransfer **f**unction), captures the audio filtering effects of a particular room from a given seat. Our room-related audio display [3] configures both the multichannel source characteristics of the sound field and the presence or ambiance (virtual room environment), as set by combination (level and timing) of direct sound, discrete **e**arlier **r**eflections ("ERs"), and later diffuse reverberation [4] [18].

This type of speaker-array approach to reverberant sound field simulation was anticipated in the work of Meyer, et. al., in the 1960s at the University of Gottingen [17]. In their system, 65 loudspeakers surrounding the listener delivered simulated reflections from the appropriate directions, and at the appropriate delay relative to the direct sound. The PSFC adds to the functionality of such a system realtime manipulation of indirect sound parameters and realtime control over the locations of virtual sound sources. Of course, only two channels of information are actually required to recreate a complete and natural-sounding spatial and temporal distribution of reflections [13]; however, in contrast to conventional two-loudspeaker sound spatialization, sound sources presented via the PSFC remain at relatively stable locations in space as listeners change the position and orientation of their heads. Additionally, such a speaker array avoids most of the difficulties associated with headphone-based presentations using binaural techniques, like front↔back reversals and problems with externalization [10] [11] [25]. The PSFC speaker array allows for a direct approach to the control of virtual source direction, resulting in soundscape stabilization that is natural and requires no head-tracking.

2. System Description

Collocated with a rear-projection stereographic (15 Hz/15 Hz left/right interleaved refresh rate, viewed through passive polarized eyewear) display[2] pictured in Fig. 2, the large (3.4 m × 8.1 m) screen and speaker array combination implement a multi-personal, spatially immersive display system [15] [22], characterized by a large field-of-view, minimized encumbrances, multiple simultaneous participants, and multimodal high-resolution displays. The stereoscopic display can be driven by an Onyx supercomputer. (As large as it is, the Synthetic World Zone's screen is dwarfed by a *really* large screen like those found in Omnimax or IMAX theaters, as suggested by Fig. 3.)

2.1 Application Program Interface

An application running on an SGI workstation connects, via a Unix socket, to the PSFC server, deployed on an HP PA-RISC workstation. This server sends commands and parameters to control sound image localization and auditory space impression via an **a**pplication **p**rogram **i**nterface. The commands and parameters are sent to a personal computer (NEC PC9800) via RS-232C, as shown by Fig. 4. A robust interface between the workstation client and the PSFC uses character-based hand-shaking transmitting 1-bit error-correcting Hamming code [28]. The protocol includes:

- DSP (**d**igital **s**ignal **p**rocessor) initialization
- sampling frequency (44.1 KHz or 48.0 KHz)
- level: source channel (#s 1, 2); volume (0–255, linear scaling)
- directionalization: source channel (#s 1, 2); azimuth ($0 \leqq \theta < 360°$); elevation ($0 \leqq \phi \leqq 90°$)
- reflection pattern index (#s 0–20)
- room size: source channel (#s 1, 2); initial time gap (0–500 [ms])
- liveness: source channel (#s 1, 2); liveness (0–255)
- pass-through mode: source channel (#s 1, 2); output speaker (#s 1–14); volume (0–255).

After interpretation by the **p**ersonal **c**omputer, control data is sent to the DSP unit via an originally-

965-8580; wlm@u-aizu.ac.jp
*7:5-601 Dept. of Electronic Engineering; Computer Acoustics and Hearing; Kogakuin Univ.; 2665-1 Nagano-machi; Hachioji-shi, Tokyo 192; ctdeneuv@sin.cc.kogakuin.ac.jp

1:www.mmc-aizu.pref.fukushima.jp/mmchome.html

2:www.mmc-aizu.pref.fukushima.jp/mmc/system/sys7.html

Proxemic Context	Display	
	audio	visual
intimate	headphones	HMD (head-mounted display)
personal	nearphones	laptop display, desktop monitor
	transaural speakers	HDTV
multipersonal or social	*speaker array*	*large-screen display*
spatially immersive displays		
public	public address	

表 1　音声，画像再生方式と視聴対象者
Table 1　Audio and Visual Displays Across Systems of Different Spatial Separation

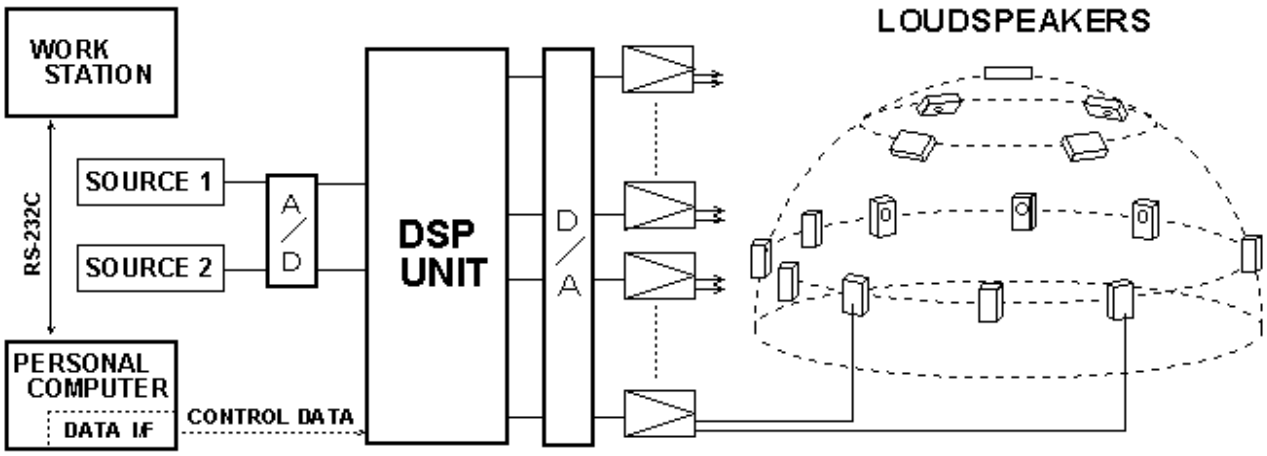

図 1　VR 音響システムの構成
Fig. 1　System Schematic

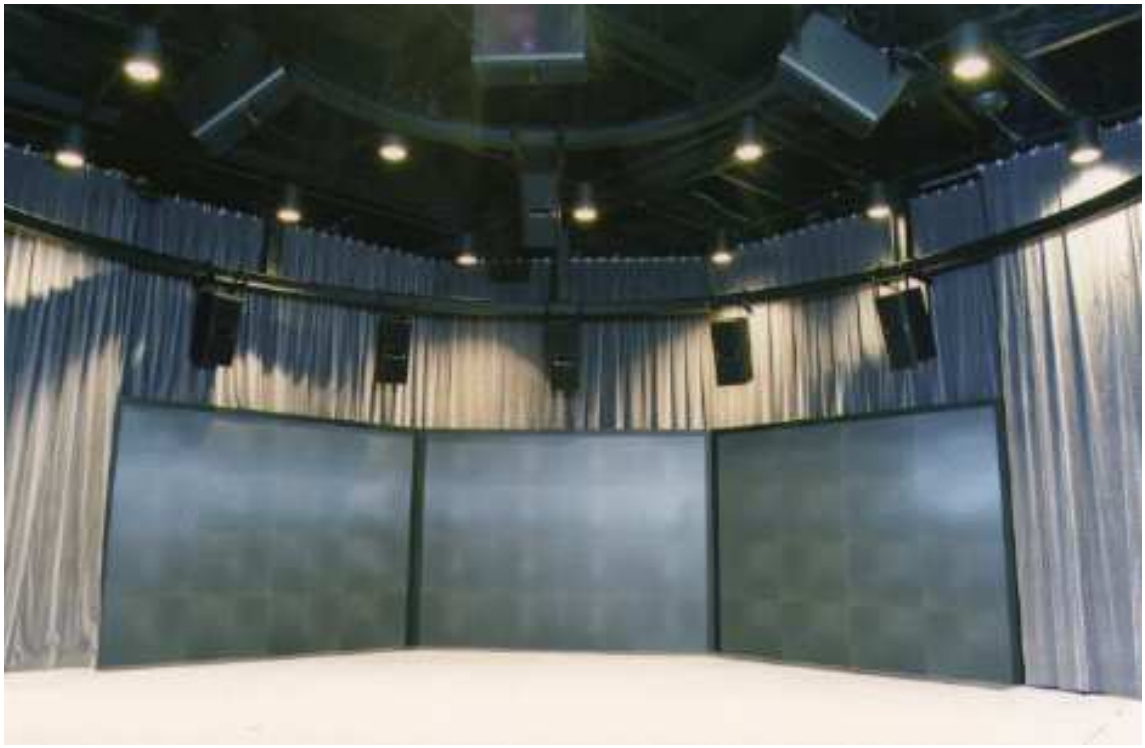

図 2　会津大学 マルチメディア センター の 人工世界 ゾーン
Fig. 2　Synthetic World Zone at the University of Aizu Multimedia Center

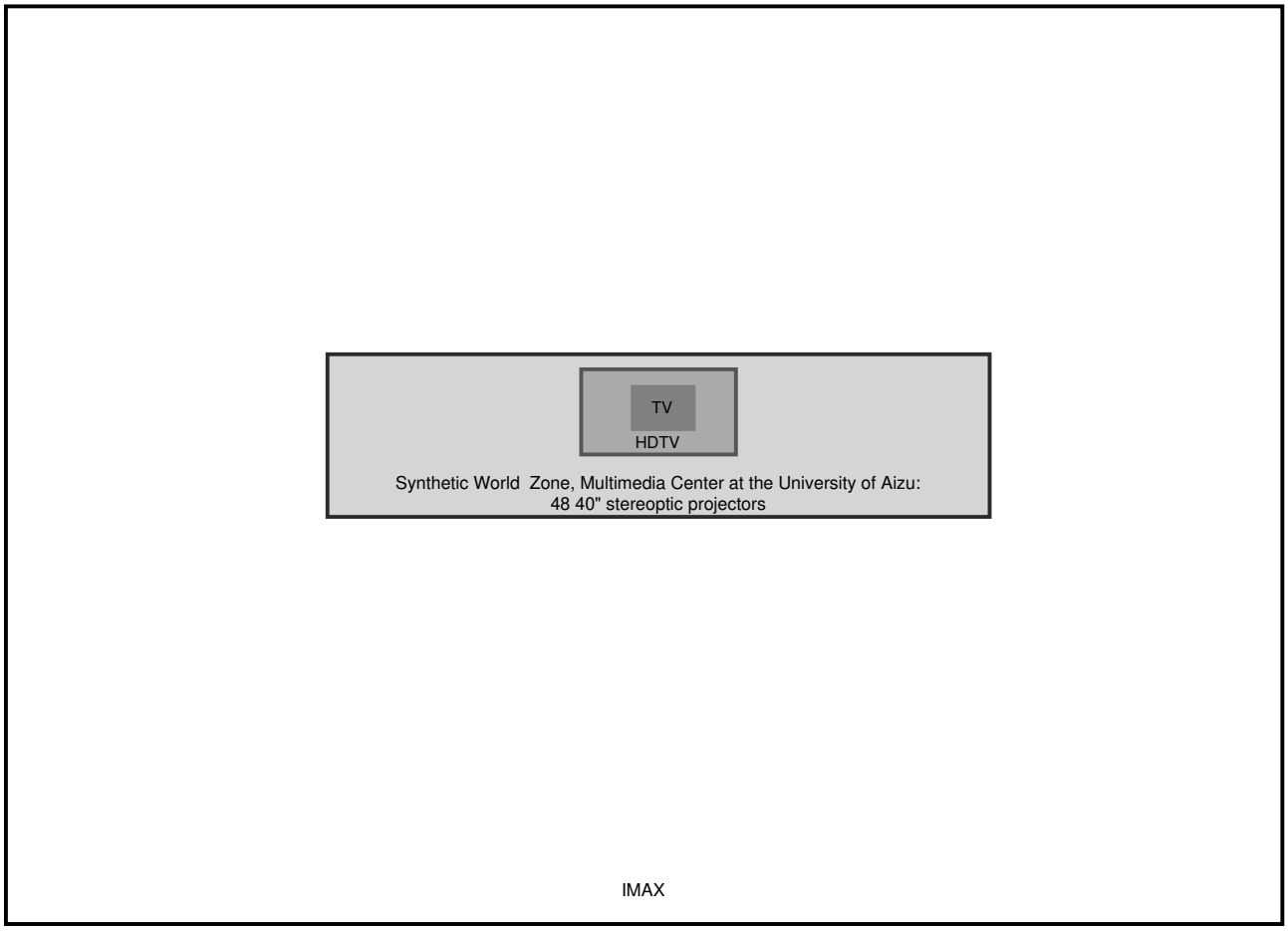

図 3 画像表示サイズの比較

Fig. 3 Screen Dimensions: comparing the Synthetic World Zone large screen display (3,500 x 10,000 pixels^2, 3.4 m × 8.1 m, 150° subtended horizontally) with conventional NTSC TV (525 lines, ∼10°), HDTV (1125 lines, ∼30°), and IMAX (18 m × 25 m, ∼180°)

designed high-speed serial interface. Two audio source channels are A-to-D converted if necessary, and fed into the DSP unit. The signals are digitally processed— delay, filtering, panning, and reverberation [2] [20]— including output channel selection and level setting, as well as synthesis and addition of indirect sounds, as described below. The signals are then D-to-A converted, and amplified through the loudspeakers in the Synthetic World Zone.

2.2 Direct Sound Directionalization

The Synthetic World Zone room is about 14 m wide. On a 10 m diameter circle of a nearly lateral plane about 3 m above the floor, ten loudspeakers are mounted; on a upper 4 m circle about 4 m above the floor, four loudspeakers are mounted; at the peak of the dome, about 4.5 m above the floor, one loudspeaker is suspended. (Because of the ceiling height in the Synthetic World Zone, the loudspeakers are not located on a sphere centered in the center of listening zone.) Each loudspeaker consists of two woofers, which drive a heavy bass sound, and one horn tweeter, which together form a spherical wave. The most distant loudspeaker can reproduce sound up to 103 dB SL.

Source direction is set by programming azimuth θ and elevation ϕ, as shown in Fig. 5, for each of two independent source channels, which can move around with an update cycle of less than 100 ms. By way of analogy to pixels and voxels, these independently directionalizable sounds are sometimes called "mixels," acronymic for sound **mix**ing **el**ements, since they are like a raster across which a soundscape is projected, defining the granularity of control. The transfer functions for the two mixels are interpolated at the audio sampling rate (and not at the slower control rate), avoiding reconstruction (aliasing) problems.

Fig. 6 shows how sound directionalization is performed. As the position of the sound image is speci-

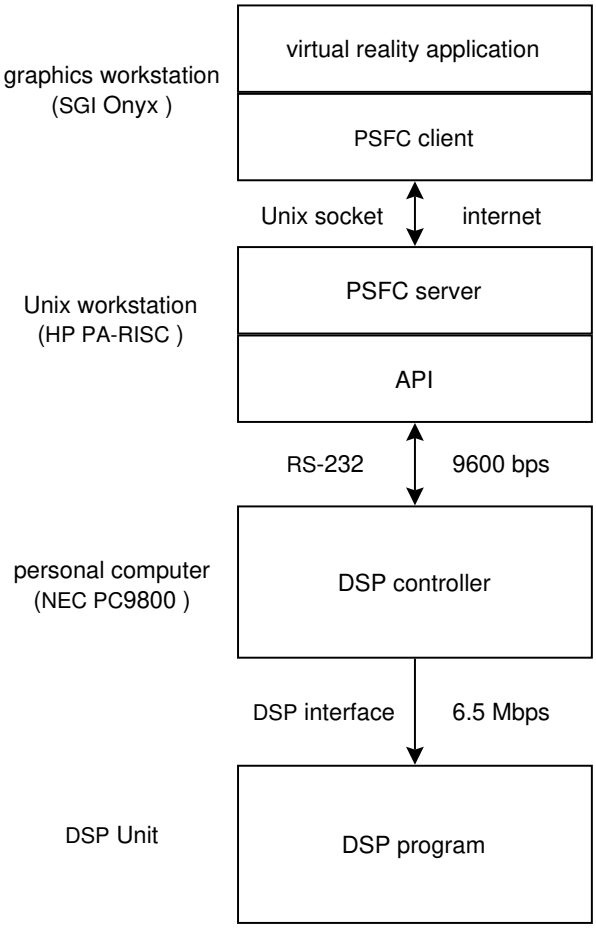

図 4　装置間インターフェースとソフトウェアのブロック構成
Fig. 4　Virtual Reality Workstation ↔ Psfc Server ↔ Dsp Controller → Dsp

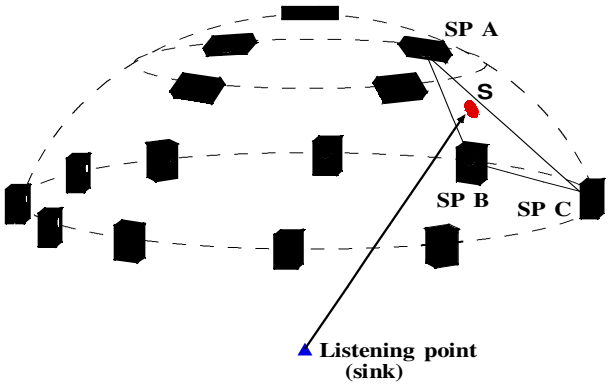

図 6　3 個のスピーカによる音像定位制御
Fig. 6　Directionalization with Loudspeakers: interpolating source image between three speakers

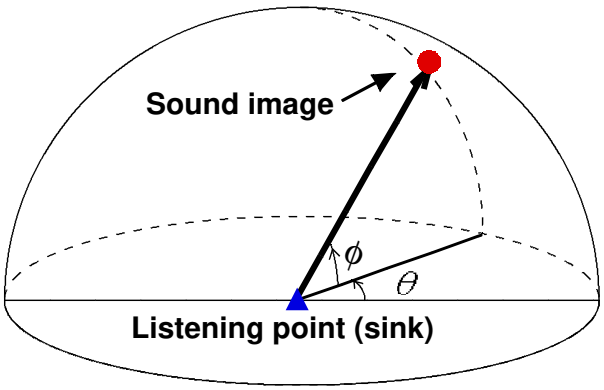

図 5　音源→受聴者: 方向制御のモデル図
Fig. 5　Source→Sink Directionalization: azimuth θ and elevation ϕ

fied by a sound directionalization command, the signal is reproduced by three loudspeakers surrounding the sound image, the level of each loudspeaker determined by the distance between the projected image source and each loudspeaker. (Such phantom images via amplitude profiling of feeds to multiple speakers recalls the Ambisonic technique [16].)

2.3 Early Reflections

Fig. 7 shows the signal flow in the DSP unit, selecting output channels and gain. The reflected sound processing blocks generate early reflections and the associated initial time gap times. The set of reflection patterns is selectable, allowing choice from among 21 contexts (seat/room combinations), currently configured for a particular listening position in Shinjuku Kousei Nenkin Kaikan (Assembly Hall), 1.5 m above the floor, about halfway back from the stage, and a little off (towards stage right) the center axis, the asymmetry yielding a richer positional sense.

Early reflection patterns were calculated via an image source model (like ray-tracing in CG, **c**omputer **g**raphics) from architectural drawings, modeling walls, ceiling and floor with characteristic reflection coefficients to obtain intensity, delay, and direction parameters [17] [13]. The FIR filters, room-related impulse responses, were generated by simulating transmission through the solid angle subtended by the respective speakers. Rather than a data-intensive, time-domain FIR filter (specifying amplitudes at a sampling rate) or a frequency-domain FIR filter (multiplying the FFTs of the sources and transfer functions), the early reflection modeling is implemented as a table of discrete (potentially sparse) delay time/normalized level pairs, the reflections being generated via a delay network.

As implemented in the PSFC, early reflections are series of pure impulses, instead of the results of convolution with a realistic wall response. In a more realistic simulation, such a wall response would include

- boundary effects, variation in frequency response, including LF (**l**ow-**f**requency) roll-off;
- material effects, absorption at both LF and HF (**h**igh-**f**requency); and
- diffusion, temporal scattering and spread of the impulse response.

To simplify the taps, reflection patterns were elided (pruned) whenever the simulated amplitude fell below a certain threshold or when the reflection closely followed another stronger (masking) one. For each of the two source signals, up to seven reflections are reproduced from each of twelve loudspeakers, for a total of 168. To correct for the greater distance between the lower ring of speakers and the center of the room compared to the upper ring, delay is added to the signals bound for the upper speakers.

ライブネス Liveness	音場効果 Effect
255	Hyperechoic space: extremely reverberant for special effects. (There is no real place with such liveness.)
254 ⋮ 128	Richly reverberant space, as in a cave, church, or tunnel.
127	Medium-sized concert hall.
126 ⋮ 52	Small-sized concert hall for chamber music.
51 ⋮ 1	Home listening room or very dead place.
0	Anechoic chamber or free field. (No echoes.)

表 2 ライブネス値とその音場効果
Table 2 Room Liveness Effects

2.4 Presence and Ambience: Room Size, Liveness, and Reverberation

The perceived sound quality of the simulated sound field is an interaction of programmed room size, liveness, and reverberation parameters, discussed in the following sections.

2.4.1 Room Size

Room size commands contribute to the perceived auditory spaciousness of the simulated environment. The apparent distance of reflective walls can be changed by varying the initial time gap (a.k.a. predelay) separating the arrival time of the direct sound from that of the first early reflected sound between 0–500 ms for each source [23] [27]. Of course, other factors, such as reverberation level, also contribute to perceived auditory spaciousness. Furthermore, the initial time gap in real spaces depends upon source location as well as room size, and so is not solely a result of changes in environmental context.

2.4.2 Liveness

Overall volume is controlled by level, but control of the simulated environment's liveness (i.e., how re-

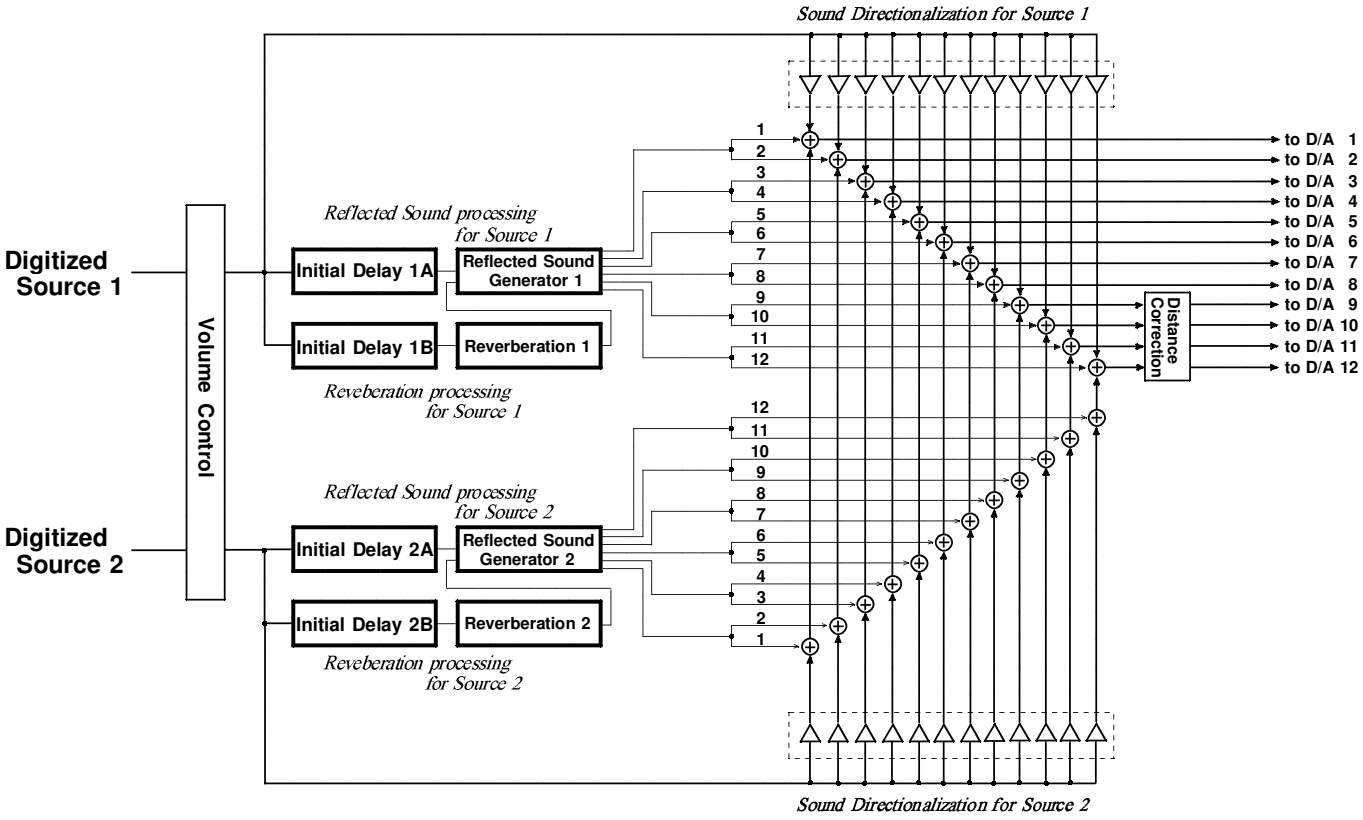

図 7　2 音源の方向制御を含めた DSP の構成

Fig. 7　Sound Processing Flow: structure of two source directional control with DSP

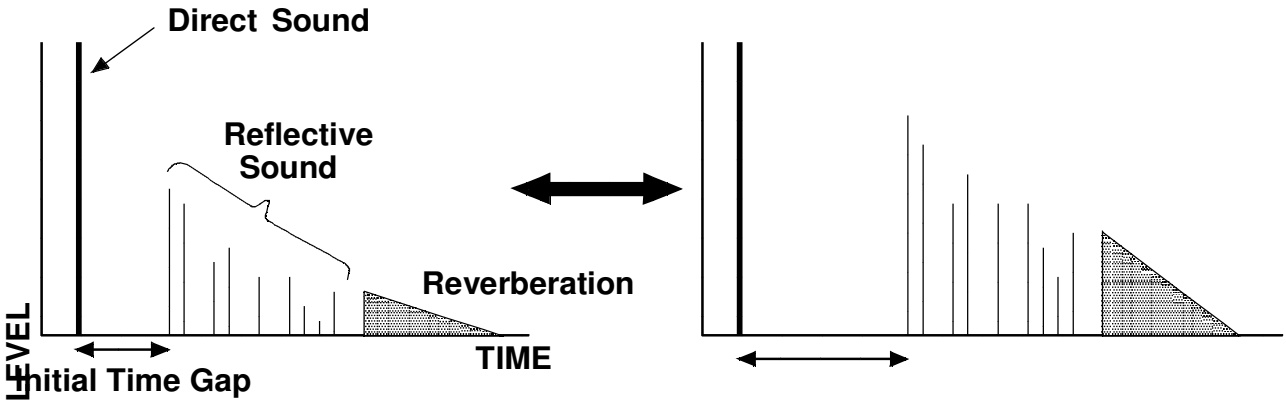

図 8　レベルと時間 (直接音、反射音、残響音): イニシャルディレイとライブネス

Fig. 8　Level vs. Time for Direct, Reflective, and Reverberant Sound: the effect on presence and indirect sound of changing specified room size and liveness (initial time gap and reflection-level decay rate). Longer predelay typically suggests a larger space, and increased liveness typically suggests a more reflective environment.

flective the walls are) is accomplished by scaling the decay rate of simluated early reflections. As shown by Table 2, the room liveness parameter is a kind of time constant, which contributes to the perceptual response characteristic termed "definition" [19]. Definition can be predicted from the ratio of the sound energy (sum of squares of coefficients) of the first 50 ms of the impulse response (including direct sound) to the total sound energy of the impulse response, as illustrated by Fig. 8. The liveness also implicitly determines the level of the following reverberation, since the reverberation level for each liveness value was prescaled to ensure smooth continuity. Changes in room size and liveness can be effected within a latency of 300 ms.

2.4.3 Reverberation

The reverberation level parameter contributes both to how spacious the simluated environment seems, and indirectly, to how distant the source seems. As discussed above, the reverberation level is scaled down if the liveness is reduced. But for a given liveness, the apparent distance of the sound source depends on the relationship been the level of the direct sound and the level of the indirect sound.

For a given virtual room simulation, varying the gain on a source relative to a fixed level of reverberant energy provides a strong cue to source distance. The distance so-cued is called the "**i**ndirect-to-**d**irect **r**atio" (IDR). In the PSFC, the IDR is controlled by simultaneously scaling the output level (which includes direct, reflective, and reverberant sound) and inversely adjusting the liveness (which includes only the indirect sound). For example, low gain and high liveness suggest a distant source, while high gain and low liveness suggest a close source.

In contrast to the early reflections, which are discretely simulated and depend upon modelled source location, reverberation is generated in a fashion that might be termed "vanilla" reverberation (avoiding the idiosyncracies of any particular solution). Exploiting the limited resolution of human hearing, the reverberation patterns are exponentially decayed noise filters, representing a Poisson distribution with time-increasing average (decreasing amplitude and increasing density), independent of source position. The reverberation processing block, detailed by Fig. 9, generates reverberations [21] and reverberation delay time, the switching point from early reflections to tail reverberations. The DSP unit generates separate reverberation for low+middle and high frequency ranges, with independently set reverberation times. Each of the two reverberation signal forks are comprised of two **g**raphic **eq**ualizers, five parallel comb filters, two all-pass filters, a 6 dB/octave low-pass filter, and a 6 dB/octave high-pass filter. By altering the delay times in the comb and all-pass filters, twelve sets of mutually incoherent reverberations are obtained (one for each primary speaker).

2.5 Pass-through Mode

A pass-through mode, useful for testing the system as well as other applications, configures null DSP with a switching matrix function by directing the destination of the two separate input signals across the 14 loudspeakers, manipulating only the direction of the virtual sound sources at an approximately fixed distance. This mode precludes automatic gain selection and reflection and reverberation generation, permitting direct patches from the input channels to the loudspeakers and allowing explicit multispeaker cross-fades.

3. Applications

Our prototype is expected to be a fertile research platform, deploying applications related to teleconferencing, virtual concerts [7], hypermedia and VR. Planned applications for the PSFC include graphical interfaces like audio windows (as shown in Fig. 10) [5], which uses a dynamic map metaphor as a multidimensional pan pot (for **pan**oramic **pot**entiometer) to configure the soundscape and VRML browsers.

For example, inspired by the cyclical nature of octaves and helical structure of a scale, we have prepared a model of a piano-style keyboard geometrically warped into a left-handed helical configuration, one octave/revolution, pitch mapped to height [12][3], generated algorithmically by Mathematica [26]. Our model is imported into (Open Inventor/)VRML, where it is driven by MIDI events, real-time or sequenced, which stream is both synthesized by a sound module and spatialized by the PSFC, so that the sound of the respective notes is directionalized with respect to sinks (Table 3), avatars of the human user, by default in the tube of the helix.

The PSFC's panoramic sound field complements the

3: www.u-aizu.ac.jp/~mcohen/spatial-media/helical-keyboard/

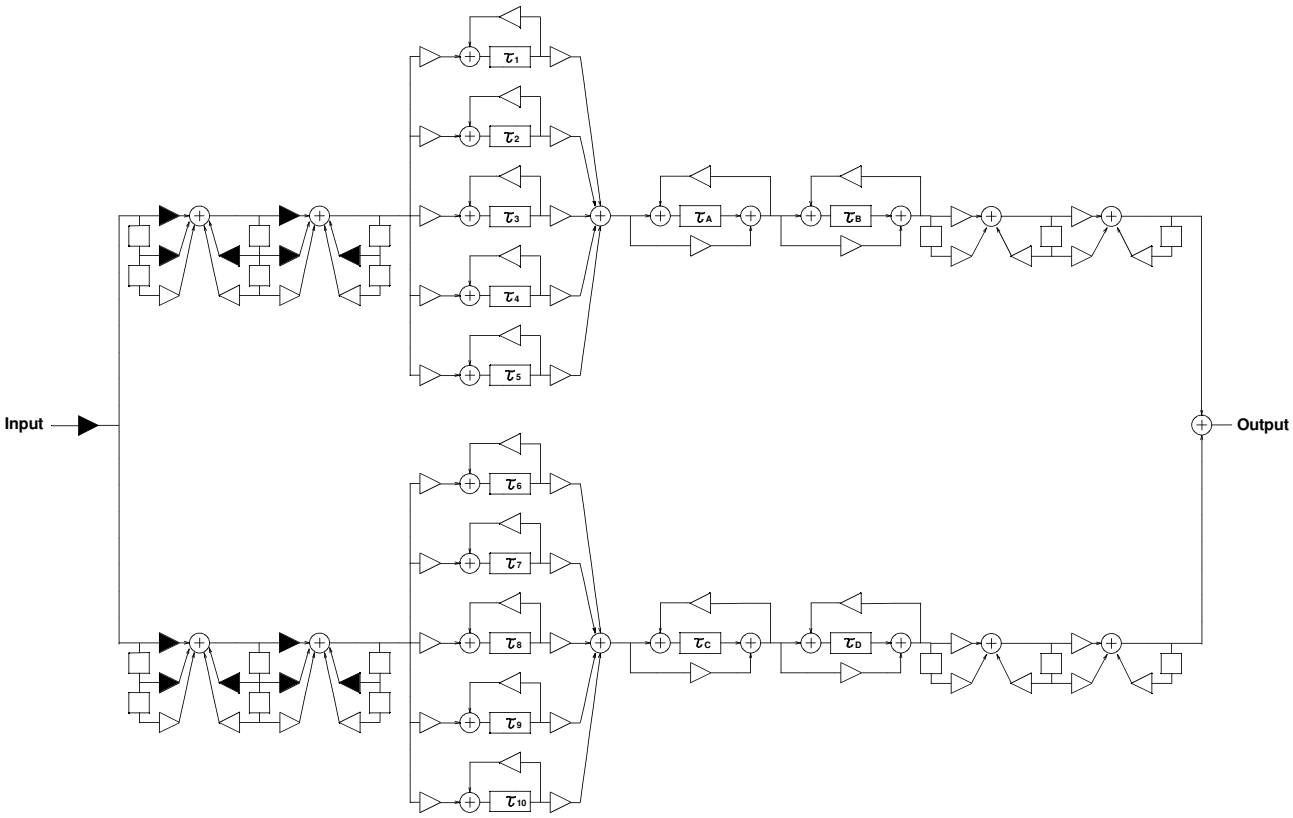

図９　残響生成ブロックダイアグラム

Fig. 9　Reverberation Processing Signal Flow: two forks (low+middle and high frequency) of two graphic equalizers, five parallel comb filters, two all-pass filters, a low-pass filter, and a high-pass filter

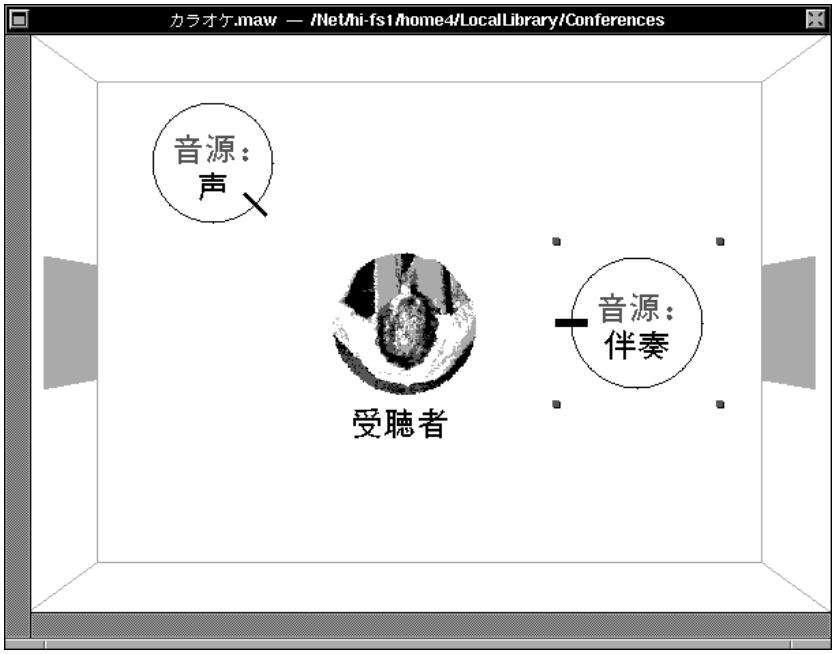

図１０　オーディオ ・ウィンドウによる仮想コンサート

Fig. 10　Virtual Concert with Audio Windows: *karaoke*-suggested separation of vocals and accompaniment

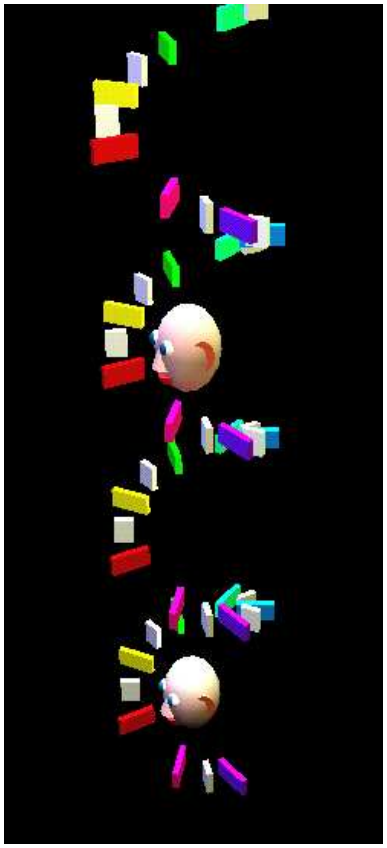

図 11　2 個所で受聴する螺旋キーボード
Fig. 11　Helical Keyboard with Two Sinks

	Role	
	音源 Source	受聴者 Sink
Function	radiation	reception
Level	amplification	sensitivity
Direction	output	input
Instance	speaker (human or loud-)	listener (human or dummy-head)

表 3　音源と受聴者
Table 3　${}^{s}\mathrm{OU}^{rce}_{Tput}$ and ${}^{s}\mathrm{IN}^{k}_{put}$

exaggerated visual display's inadequacies for immersion, since flying into the helix from the outside envelopes the listener as it overflows the visual system's field-of-view. The helical keyboard's multiple sinks are natural idioms for the PSFC, compensating for the fact that the speakers span the ceiling but not the floor; multiple sinks can be used to normalize octave (ex.: one sink for harmony and one for melody, as shown in Fig. 11), so that all sound images are in the upper hemisphere [8] [6].

4. Limitations

The PSFC has several significant restrictions. Perhaps the most direct problem is the lack of omnidirectionality due to the base of the speaker semidome being elevated; as the speakers are all well ($\geq$3 m) above head level, the range of possible sound directions (display space) is unnecessarily limited. Even though the elevation control parameter ranges [0°, 90°], the geometry of the display actually ranges [$\sim$18.6°, 90°]. The Synthetic World Zone has floor-mounted speakers, but these are not integrated with the PSFC, nor is the speaker on the (overhead) pole.

There is an additional problem that the spatial and temporal patterns of simulated reflections generated by the system are based upon an "ideal" listening position. For listeners sitting at increasing distances from the "ideal" position, the spatial and temporal pattern of simulated reflections that each hears is progressively further from those described by the architectural model for which the pattern was generated. A more subtle problem is that the reflections are not diffused, since they are implemented as simple <delay, normalized level> tables. Echos sound like unnatural "slaps" off perfectly reflective walls, rather than warmer-sounding reflections that result from the inclusion of boundary effects due to the size and shape of reflecting surfaces.

Also, the PSFC has no direct control of time delays, which are critical for directional cueing. Since the PSFC doesn't simulate interaural cues directly, relying on speaker location to implicitly generate the ITDs, left–right positioning is much weaker than it could be. When a multichannel loudspeaker system is used for more than a few listeners, the problem of ensuring a common experience for all is addressed better by modulating ITD than by modulating level. Such an feature would also extend the range of sound

source directions that the system could produce, compensating for the speaker placement restriction on attainable image angles. A consequence of the lack of direct manipulation of ITD is the inability to generate the kind of interaural pitch shift that is normally associated with sound sources that move rapidly about the listener's head (since a shifting delay always produces a shifting pitch).

Another serious liability is the lack of direct control over apparent distance. Since the loudspeakers are all approximately at a constant distance from the room's center, apparent distance can only be suggested by varying the direct:reverberant ratio. Furthermore, the system's linear gain adjustment causes the range to be oversampled at high levels and undersampled at low levels. This means that calibration of distance cues becomes very difficult, as the adjustment to keep reverb level constant is too coarse at the low end.

We have neither objectively validated the physical simulation capability of the PSFC nor subjectively measured its psychoacoustic resolution. Presumably such experiences would influence and be influenced by visual stimuli, which would affect calibration attempts.

5. Conclusion

The PSFC can configure both the context and the content of a virtual sound field: the context is the ambiance or presence— the room size, liveness, reflection pattern, and overall level; the content is the source direction— azimuth and elevation and suggested distance. The direct sound moves by amplitude panning; the reflected sound moves by rotating the impulse response; the reverberant sound is orientation-independent.

The hemispherical speaker array's audio presentation is complemented by the wide-screen visual presentation. The spatially immersive environment is consistent with the aims of "calm technology" [24], providing a natural group experience, wide visual field-of-view, and comfort (no fatigue-inducing or cumbersome head-mounted display). Such a system can be described as "roomware," software for a room as a component of a cooperative building, putting the users *inside* the computer. This approach is also related to the idea of an "immobot," an immobile robot that concentrates on attending and servicing the needs of colocated human users (rather than the traditional robotic tasks of exploration and manipulation of an external environment).

We plan to use the PSFC not only for practical applications like teleconferencing, but also for more artistic expressions, sculpting and shaping sound as it speeds through space. For example, we hope to use the direct patch matrix to explore non-conventional types of spatial sound, like extensions to [2, p. 244]'s idea of bipolar sources spinning synchronously at opposite ends of a room's diameter.

Acknowledgements

The SGI↔PSFC interface was implemented by Kimitaka Ishikawa (石川 君孝) and Minefumi Hirose (広瀬 峰文). This research was supported in part by grants from the Fukushima Prefectural Foundation for the Advancement of Science and Education and the NTT Human Interface Laboratories.

参 考 文 献

[1] Katsumi Amano, Fumio Matsushita, Hirofumi Yanagawa, Michael Cohen, Jens Herder, Yoshiharu Koba, and Mikio Tohyama. PSFC: the **P**ioneer **S**ound **F**ield **C**ontrol System at the University of Aizu Multimedia Center. In *Proc. Ro-Man: 5^{th} IEEE Int. Workshop on* **Ro***bot and Hu*man *Communication*, pages 495–499, Tsukuba, Japan, November 1996. ISBN 0-7803-3253-9.

[2] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, 1994. ISBN 0-12-084735-3.

[3] Jens Blauert. Acoustical simulation and auralization for VR and other applications. In *Proc. ASVA: Int. Symposium on* **S***imulation,* **V***isualization and* **A***uralization for* **A***coustic Research and Education*, pages 261–268, Tokyo, April 1997.

[4] John M. Chowning. The simulation of moving sound sources. *Computer Music Journal*, 1(3):48–52, June 1977.

[5] Michael Cohen. Integrating graphical and audio windows. *Presence: Teleoperators and Virtual Environments*, 1(4):468–481, Fall 1993. ISSN 1054-7460.

[6] Michael Cohen. Quantity of presence: Beyond person, number, and pronouns. In Tosiyasu L. Kunii and A. Luciani, editors, *Cyberworlds*, chapter 19, pages 267–286. Springer Verlag, 1998. ISBN 4-431-70207-5.

[7] Michael Cohen and Nobuo Koizumi. Audio Windows for Virtual Concerts I. In Susumu Tachi, editor, *Video Proc. ICAT/VRST:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence/Conf. on* **V***irtual* **R***eality* **S***oftware and* **T***echnology*, page 253, Makuhari, Chiba; Japan, November 1995. ACM-SIGCHI (TBD), SICE (**S**ociety of **I**nstrument and **C**ontrol **E**ngineers), JTTAS (**J**apan **T**echnology **T**ransfer **A**ssociation), and NIKKEI

(**Ni**hon **Kei**zai Shimbun, Inc.).

[8] Michael Cohen and Nobuo Koizumi. Virtual gain for audio windows. *Presence: Teleoperators and Virtual Environments*, 7(1), February 1998. ISSN 1054-7460.

[9] C. Cruz-Neira, D. Sandin, and T. DeFanti. Surround-Screen Projection-Based Virtual Reality: Design and Implementation of the CAVE. In SIGGRAPH *Proc.*, pages 135–142, 1993.

[10] N. I. Durlach, A. Rigpulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel. On the externalization of auditory images. *Presence: Teleoperators and Virtual Environments*, 1(2):251–257, Spring 1992. ISSN 1054-7460.

[11] Scott H. Foster, Elizabeth M. Wenzel, and R. Michael Taylor. Real-time synthesis of complex acoustic environments. In *Proc.* (IEEE) *ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October 1991. Summary.

[12] Jens Herder and Michael Cohen. Project report: Design of a helical keyboard. In *Proc. ICAD:* **I***nt.* **C***onf.* **A***uditory* **D***isplay*, pages 139–142, Palo Alto, CA, November 1996. `www.santafe.edu/~icad/ICAD96/proc96/herder.htm`.

[13] Gary Kendall and William L. Martens. Spatial reverberation. U.S. Patent 4,731,848; European patent specification 0 207 084 B1, 1988.

[14] Myron W. Krueger. *Artificial Reality II.* Addison-Wesley, Reading, MA, 1991. 0-201-52260-8.

[15] Ed Lantz, Steve Bryson, David Zeltzer, Mark T. Bolas, Bertrand de La Chappelle, and David Bennett. The future of virtual reality: Head mounted displays versus spatially immersive displays (panel). In SIGGRAPH *Proc.*, pages 485–486, New Orleans, LA, August 1996.

[16] David G. Malham and Anthony Myatt. 3-D Sound Spatialization using Ambisonic Techniques. *Computer Music Journal*, 19(4):58–70, Winter 1995.

[17] E. Meyer, W. Bugtorf, and P. Damaske. An apparatus for electroacoustical simulation of sound fields: Subjective auditory effects at the transition between coherence and incoherence. *Acustica*, 15:339–344, 1965.

[18] F. Richard Moore. A general model for spatial processing of sounds. *Computer Music Journal*, 7(3):6–15, Fall 1983.

[19] R.A. Rasch and R. Plomp. The listener and the acoustic environment. In D. Deutsch, editor, *The Psychology of Music*, pages 135–147. Academic Press, 1984. ISBN 0-12-213560-1 or 0-12-213562-8.

[20] Curtis Roads. *The Computer Music Tutorial.* MIT Press, 1996. ISBN 0-252-18158-4.

[21] Manfred R. Schroeder. Natural sounding artificial revereration. *J.* **A***ud.* **E***ng.* **S***oc.*, 10(3):219–223, 1962.

[22] Kishore Swaminathan and Steve Sato. Interaction design for large displays. *interactions*, pages 15–24, January/February 1997.

[23] Mikio Tohyama, Hideo Suzuki, and Yoichi Ando. *The Nature and Technology of Acoustic Space.* Academic Press, London, 1995. ISBN 0-12-692590-9.

[24] Mark Weiser and John Seely Brown. Designing calm technology. *PowerGrid Journal*, 1.01, July 1996. `http://powergrid.electricit.com/1.01`.

[25] Elizabeth M. Wenzel. Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, 1(1):80–107, 1992. ISSN 1054-7460.

[26] Stephen Wolfram. *The Mathematica Book.* Wolfram Media/Cambridge University Press, third edition, 1996. ISBN 0-9650532-0-2.

[27] パイオニア (株) 大森技術研究所. マルチスピーカ方式の音場創成システム. 日本音響学会誌, 48(4):275–278, 1992.

[28] 木場 美治. マルチスピーカによる VR 音響システムについて. Master's thesis, 工学院大学, 1996. 工学部.

(1997 年 7 月 7 日受付)