SCAPE SYNTHESIS: GENERATION OF A NOVEL ACOUSTICAL SPACE BASED ON AUDITORY PERCEPTUAL FEATURES

Shoichi Koyama
The University of Tokyo
Information Science and Technology

Risa Suzuki
The University of Tokyo
Interdisciplinary Information Studies

ABSTRACT

The human auditory system can organize environmental sounds into perceptually meaningful elements. This process is called "Auditory Scene Analysis"[1]. Our sound installation, "Scape Synthesis", aimed to generate a unprecedented acoustical space and provide a novel perceptual experience of hearing by modulating the process of sound organization. The installation was exhibited at The University of Tokyo, "iii Exhibition 10" (2008).

1. INTRODUCTION

Humans can recognize individual surrounding sounds even in a noisy environment. For example, the ability to focus one's listening attention on a single speaker among a mixture of conversations is well known (cocktail party effect)[2, 6], and when listening to music, we can easily segregate an instrument sound from the mixed sound of various instruments. The separation of superimposed signals is ill-posed problem. Humans can perceptually solve this problem in real-time. Albert Bregman, a psychologist, named these perceptual processes as "Auditory Scene Analysis" in 1990[1]. Surrounding sounds are segregated and/or integrated based on some regularity conditions[4].

1. Unrelated sounds seldom start or stop at exactly the same time.

2. A single sound or a sequence of sounds from the same source tends to change its properties smoothly and slowly.

3. Vibrations give rise to an acoustic pattern in which the frequency components are multiples of a common fundamental.

4. Many changes will affect all the components of the resulting sound in the same way and at the same time.

This process can be related to principles of the Gestalt psychology.

Now, if key elements of auditory stream segregation are gradually broken, what happens? Our installation presents the process of spatial collapse and reconstruction of segregated sounds. Some installations and art works have been made to present the spatial decomposition of sounds[7, 8, 5]. However, there is very few works that are focused on auditory perceptual features such as auditory scene analysis. The process of our work is intended to generate a novel acoustical space and provide a novel perceptual experience of hearing.

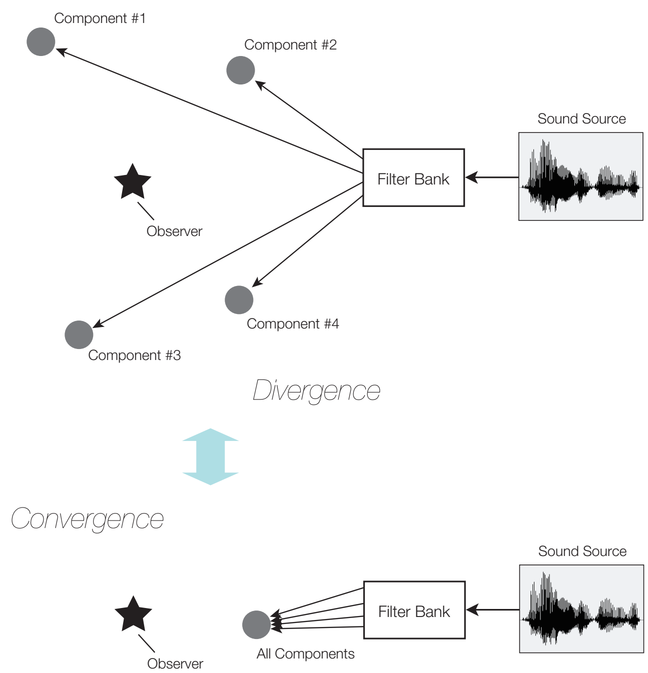

Figure 1. Sounds are decomposed to some frequency components and repeat spatial divergence and convergence.

2. "SCAPE SYNTHESIS"

2.1. Divergence / Convergence of Sounds

In our work, once the harmonic structures of sounds are broken by disassembling the sound into frequency components,

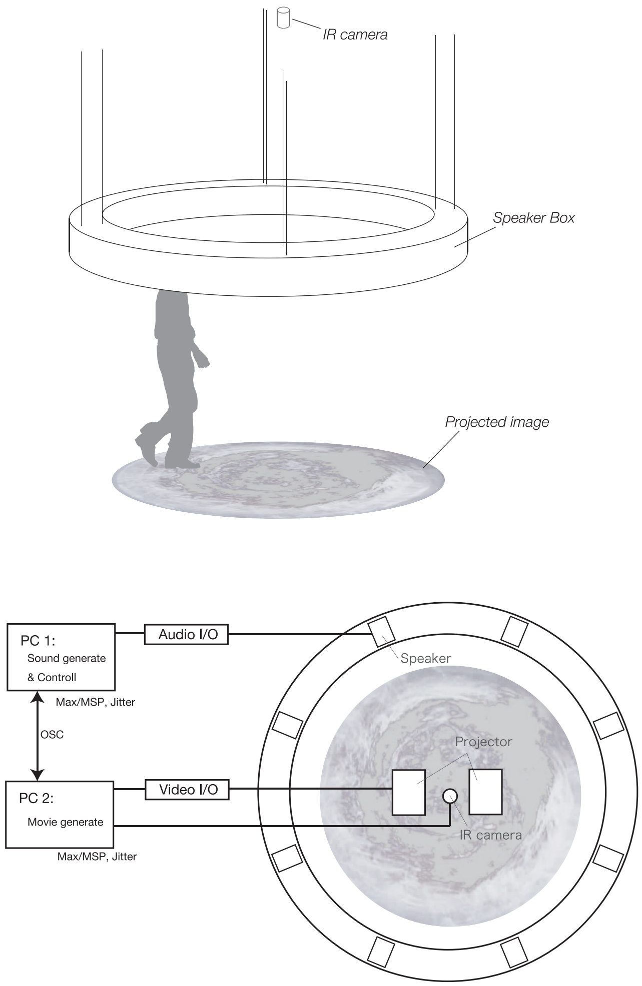

Figure 2. System diagram.

and then the frequency components are separately prsented in space by different speakers, which surround the observer. In other words, Gestalt of hearing is segregated in temporal domain and integrated in space. These components repeat spatial scattering and gathering. Therefore, sounds change gradually between disassembling and reconstructing.

We call this perceptual disassembling and reconstructing of sounds "divergence/convergence of sounds" (see Fig.1). As the "divergence/convergence of sounds" can be influenced by the feature of sound, the presented sounds changed in stable and unstable auditory scene, a concrete sound changed to abstract sounds, and vice versa. In addition to the change of the sound itself, the sound components moved dynamically in space.

2.2. System

Figure 2 shows system diagram of this work. We used 8ch surround speaker system (ES6300, NetCIRA by FOSTEX,

Figure 3. A photo of "scape synthesis".

Japan). Audio signals are transfered to each speaker by Ethernet cable[3]. Sounds were disassembled to frequency components by FFT based filter bank using Max/MSP.

Sound components are allocated by panning. The volume balance of speakers determines each location of sound component. The location and movement of sound components are also controlled using Max/MSP.

To visually support auditory perception, we projected prerecorded images that moved in the same direction. The images were of fluid diffusion, which were generated using Jitter. At divergence of sounds, movies of fluid diffusion were played. At convergence of sounds, movies played in backwards. The movement of sound components started when a observer stood at the center of this work. An infrared camera was used for detecting the observer.

Figure 3 shows a photo of this work. We made the speaker box in the shape of circular ring to hide the speakers. The size of speaker box is 4m in diameter, and 0.15m in height. The 8 speakers were placed with the same angular interval in this box. This ring speaker box was hung from the ceiling. Fluid images were projected onto the floor just under the ring.

2.3. Sequence

Figure 4 shows the sequence of observer's experience. Left figures are projected images of fluid diffusion and right figures show the movement of sound components. We used three audio samples and each sample was disassembled into 8 frequency components, resulting in 24 components moving around the listener.

First, sound components were scattered and swirled, which coincided with projected images (Fig. 4A). When a observer stood at the center of projected images, the observer was detected. The sound components started to converge and reconstruct original sounds (Fig. 4B-C). At this time,

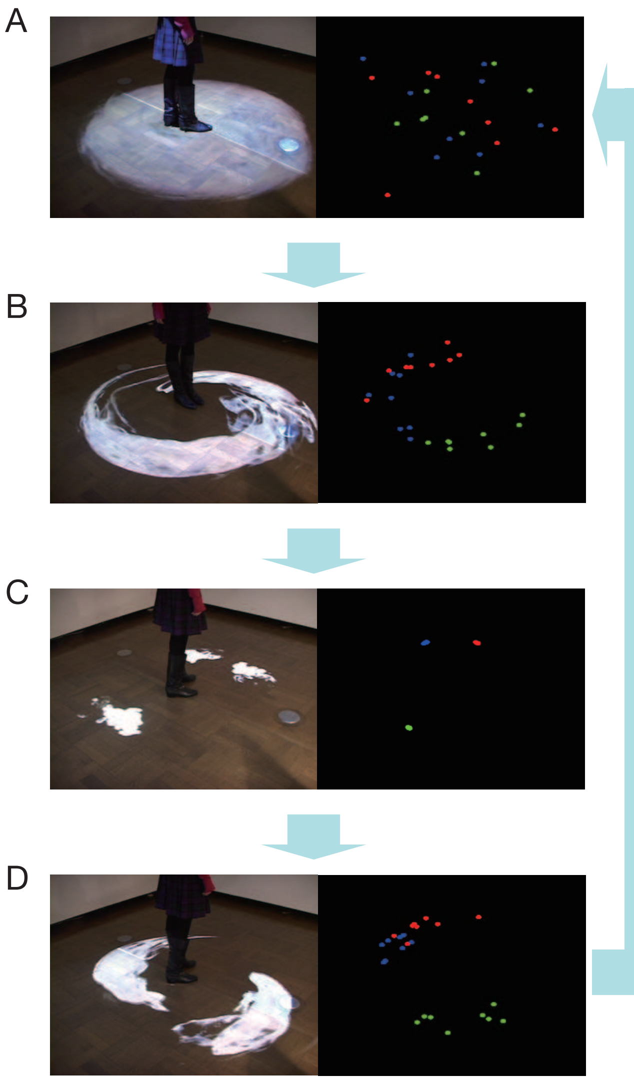

Figure 4. The process of "scape synthesis".

three sound sources were located in the direction as projected images. After a while, sounds started to diverge and the components are spatially scattered (Fig. 4C-D). This process takes about one minute and repeated with several sounds.

We used human voices and many physical sounds, such as sounds of opening doors, footsteps, tearing up paper, pouring water, playing guitar, and more. These sounds were effective because they had rich harmonic sounds.

2.4. Exhibition

"Scape Synthesis" was exhibited at The University of Tokyo, "iii Exhibition 10" in December 2008 (see Fig.5). More than 500 people visited this exhibition and experienced this work. Many people felt that it was an unprecedented experience of auditory uncertainty and settling. And that also led psychological uncertainty and settling.

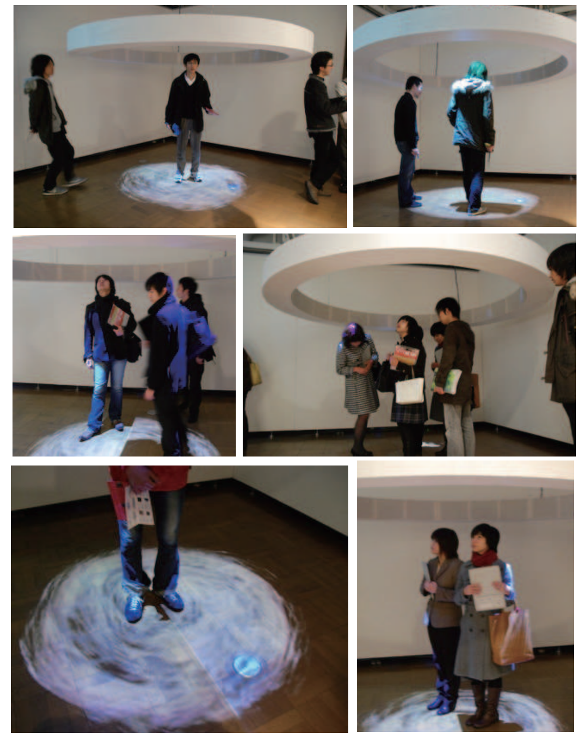

Figure 5. Photos of "scape synthesis" at the exhibition.

3. CONCLUSION

Several works have been made to present modifying acoustic space. But very few attempts have been focused on properties of auditory perception. In "Scape Synthesis", we presented sequential change between stable and unstable acoustic space based on auditory scene analysis. This process is perception of spatial Gestalt psychology and observers feel changes between psychological uncertainty and settling. It is possible to use the method of wave field synthesis (WFS) to allocate sound components for high realistic sensation. What is more, using a technique of three dimensional acoustic field synthesis, this process would be more effective.

4. REFERENCES

[1] A. S. Bregman, *Auditory Scene Analysis: Perceptual Organization of Sound*. Bradford Books, 1990.

[2] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *Journal of Acoustic Society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[3] EtherSound, http://www.ethersound.com/.

[4] S. McAdams and E. Bigand, *Thinking in Sound - The Cognitive Psychology of Human Audition*. OXFORD SCIENCE PUBLICATIONS, 1993.

[5] A. McLeran, C. Roads, B. L. Sturm, and J. J. Shynk, "Granular sound spatialization using dictionary-based methods," *5th Sound and Music Computing Conf.*, 2008.

[6] B. C. J. Moore, *Introduction to the Psychology of Hearing*. ACADEMIC PRESS, 2003.

[7] C. Roads, *Microsound*. MIT Press, 2004.

[8] I. Xenakis, "Elements of stochastic music," *Gravensaner Blatter*, vol. 18, pp. 84–105, 1960.