Proceedings of the Second IASTED International Conference
HUMAN-COMPUTER INTERACTION
March 14–16, 2007, Chamonix, France
ISBN Hardcopy: 978-0-88986-654-6 / CD: 978-0-88986-655-3

DESIGNING AN AURAL USER INTERFACE FOR ENHANCING SPATIAL CONCEPTUALIZATION

Shigueo Nomura, Takuya Utsunomiya, Masayoshi Tsuchinaga, Takayuki Shiose, Hiroshi Kawakami, Osamu Katai
Department of Systems Science, Graduate School of Informatics
Kyoto University
Kyoto 606-8501 Japan
email: {shigueo, utsunomiya, tsuchinaga}@sys.i.kyoto-u.ac.jp
email: {shiose, kawakami, katai}@i.kyoto-u.ac.jp
Keiji Yamanaka
Faculty of Electrical Engineering
Federal University of Uberlândia
Uberlândia 38400-902 Brasil
email: keiji@ufu.br

ABSTRACT
We propose an approach for enhancing the spatial conceptualization performance by subjects traveling along different types of virtual tracks in experiments via virtual 3D acoustic space system. Subjects in a navigation task must perceive alterations on aural surfaces before categorizing these virtual tracks. We adopted natural-sounding tones as components of these aural surfaces. Fourteen subjects participated in these experiments. First evaluation results revealed that the proposed approach improved the spatial conceptualization performance by subjects. We concluded that the approach can be an essential requirement for designing novel aural user interfaces as supporting systems for visually impaired and elderly people.

KEY WORDS
User Interface, Spatial Conceptualization, Visually Impaired, Aural Surface, Virtual Track, 3D Acoustic Space.

1. Introduction

It is known that the auditory system has been considered a major tool of perception employed by the visually impaired because it provides rich information about the world around us.

In this context, everyday listening [1] [2] is the experience of listening to events rather than sounds and makes important role for acquiring spatial information around us.

Basically, we are concerned with listening to the things going on around us and hearing which things are important to avoid and which might offer possibilities for action.

Traditional audition approaches have ignored such experience that seems qualitatively different from listening to music (perceptual dimensions of the sound itself).

We have found that everyday listening skills have been virtually ignored in interaction with computers or in traditional interfaces for supporting the visually impaired.

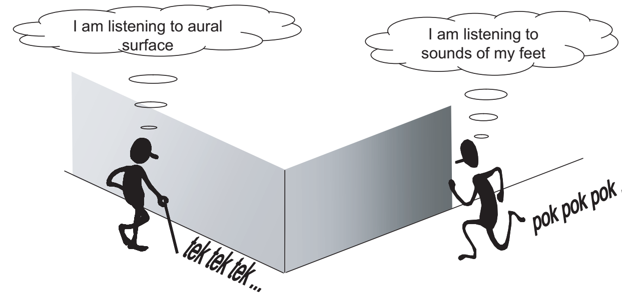

Figure 1. Visually impaired experience of listening to events (aural surfaces) rather than sounds.

We believe that a comprehensive account of everyday listening is possible to emerge.

Ito [3] investigated how the events in everyday listening are perceived by blind and sighted pedestrians in a navigation task. The task consisted of walking an underground city and describing the perceived sounds. The investigation results lead to conclusion that the visually impaired subjects were more able to perceive the sounds as events (aural surfaces) than sighted subjects. Figure 1 illustrates a difference of visually impaired persons in the experience of listening to events rather than sounds. We can verify that the blind pedestrian perceives aural surfaces as events rather than sounds.

According to Donker's investigation [4], the blind users do not want to know what an object exactly looks like, but they want to understand the object structure. Our approach is to give access to the essential layout structures such as aural surfaces of objects interacting with these users.

By taking advantage of this experience of listening to events rather than sounds, we believe that the user may utilize the same skills employed in such everyday tasks as crossing the street. An example of such skills is our "built-up association" of reverberations with empty space. If all other things are equal, the more reverberance in a room means that there is more space. So this built-up association should provide a spatial conceptualization based on user experience and familiarization with everyday listening. Then, it would provide user-friendly interface for acquiring spatial information without hard cross-modal training.

In our previous work [5], we proposed to break away from typical use of "artificial-sounding" tones like in vOICe Learning Edition [6] toward novel interfaces based on "natural-sounding" tones that consider everyday listening.

Erulkar [7] says that the ability to localize a sound in natural space is present in nearly animals that posses a hearing mechanism. Also, accurate sound localization is ecologically important for most animal species including human beings since it is fundamental for survival, and interaction with the world around them [8].

In this work, we propose an approach for enhancing the spatial conceptualization performance by subjects in experiments using a virtual 3D acoustic space system. The navigation task for subjects is to travel along different types of virtual tracks and perceive alterations on aural surfaces for categorizing these tracks.

By considering that ecological importance of accurate sound localization, the main objective of this work is to verify the influence of reversing the traveling direction on the spatial conceptualization performance by subjects.

To create the virtual 3D acoustic space for experiments, we based on apparatus utilized in our previously investigated topics such as "sound visualization [9]," and "perception of crossability [10]."

Section 2 presents the proposed approach for designing nonspeech sound, implementing 3D sound apparatus, creating 3D virtual tracks, and evaluating spatial conceptualization performance. First evaluation results are presented in Section 3. In Section 4, the spatial conceptualization performance by subjects traveling along the 3D virtual tracks is discussed. Section 5 presents a conclusion.

2. Proposed Approach

The proposed approach consists of designing nonspeech sound, implementing 3D sound apparatus, creating 3D virtual tracks, and evaluating spatial conceptualization performance. The evaluation process is to investigate spatial conceptualization performance by subjects in experiments using the designed aural interface. Also, the investigation consists of verifying the viability and usability of this design.

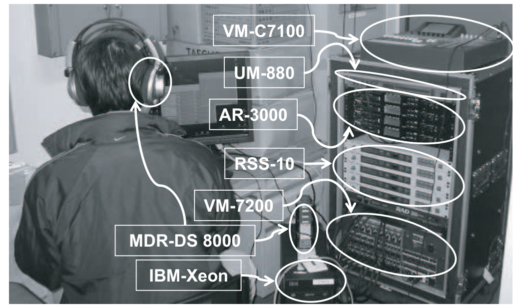

Figure 2. Front view of devices for generating nonspeech tones based on previous work [5].

2.1 Designing Nonspeech Sound

Firstly, we designed and evaluated "artificial-sounding" and "natural-sounding" tones as nonspeech sounds (spatialized sound source) for aural interface users conceptualizing spatial information. The evaluation showed that natural-sounding tones work better than artificial-sounding ones to categorize different types of 3D virtual acoustic tracks. The moving direction of subjects that participated in the experiments was forward, that is, they heard sounds located ahead of their ears [5].

Since the suitable nonspeech sound as the result of that evaluation was the natural-sounding tone, we adopted this tone to get the performance results of subjects evaluated by experiments in this work. A virtual 3D acoustic space system described below adds effects such as reverberation and reflection into sound source.

The nonspeech sound is designed and spatialized as follows:

- The sound source is the everyday world sound of fan noise captured by precise microphones [11].
- The effects are the appropriate reverberation and reflection levels resulted as -30 dB in our previous work [10].

2.2 Setting 3D Sound Apparatus

We set the virtual 3D acoustic space system based on sound space processors as the scene shown in Figure 2. Figure 3 shows the schematic overview of the acoustic space system as apparatus for the experiments. The overview provides a chart showing the audio movement and linking scheme of devices. The devices are identified by their model numbers. The system generates the natural-sounding tone as nonspeech sound previously designed and creates 3D virtual acoustic tracks for subjects traveling along these tracks during the evaluation process.

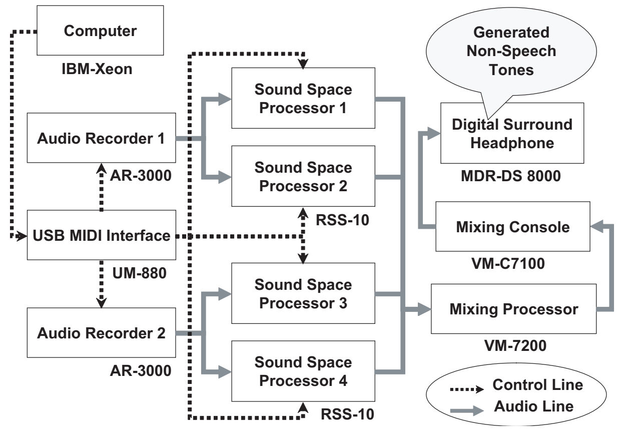

Figure 3. Schematic overview of apparatus for experiments based on previous work [5].

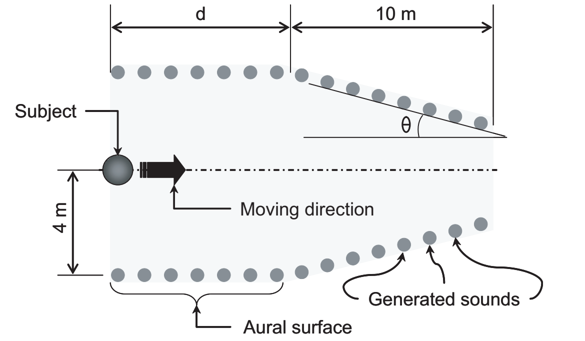

Figure 4. Upper view of a 3D virtual track for experiments.

2.3 Creating 3D Virtual Tracks

A 3D virtual track is constituted by aural surfaces representing acoustic walls as shown in Figure 4. The nonspeech sounds generated by the 3D acoustic space system form the aural surfaces of a virtual track. A typical 3D virtual track as shown in Figure 4 is designed as follows:

1. The distance d is calculated by equation $d = v.t$, where v is $4\ km/h$ and $t \in \{3.0, 3.5, 4.0, \ldots, 6.0\}s$. This distance corresponds to a straight part (without de□ection in the aural surface) of the virtual track. On the other hand, the distance of $10\ m$ corresponds to the part where the aural surface can deflect for the track categorization task by subjects.

2. We have different geometric shapes of virtual tracks according to the aural surface deflection measured by the angle θ in Figure 4, as follows:

   - Convergent track for $\theta \in \{3^o, 5^o, 30^o\}$. The deflection in the aural surface causes a gradually narrower track for subjects traveling along this environment.
   - Straight track for $\theta = 0$. There is no deflection in the aural surface for subjects traveling along the entire track.
   - Divergent track for $\theta \in \{-3^o, -5^o, -30^o\}$. The reversed deflection in relation to the above convergent track causes a gradually wider track for subjects traveling along this environment.

3. Four nonspeech sounds (natural-sounding tones) generated by the 3D virtual acoustic space system form a unity of the aural surface from each virtual track.

2.4 Evaluating Spatial Conceptualization Performance

Basically, experiments consist of evaluating the effect on spatial conceptualization performance after reversing the moving direction of subjects traveling along the different virtual tracks. Fourteen subjects participated in these experiments. Participants had no audition problems. No visual information was necessary for trying a navigation task. Each trial for subjects in navigation tasks was set as follows:

1. Each time t was randomly chosen for each trial to calculate the distance d in Figure 4. Then, the subject travels along different length of track in each trial. This strategy is to avoid learning process by subjects and increase the confidence level of experimental results.

2. A 3D virtual track suffers deflections in its aural surfaces after the distance d as shown in Figure 4. After starting this deflection, the time is counted until the subject presses a key on the keyboard as perception response. This time works as reference to validate the experimental data. If the reference time is equal or less than zero, then it means that the subject pressed a decision key without certainty. In this case, the corresponding result was discarded.

In each trial, the subject's task is to perceive deflection or no deflection in the aural surface and categorize the traveled track performing the following decisions:

- When the subject perceives that the traveled track is converging as shown in Figure 5a, then he/she must press the key corresponding to "←" on the keyboard.
- When the subject perceives no deflection in the aural surfaces of the traveled track as shown in Figure 5b, then he/she must press the key corresponding to "↓" on the keyboard.
- When the subject perceives that the traveled track is diverging as shown in Figure 5c, then he/she must press the key corresponding to "→" on the keyboard.

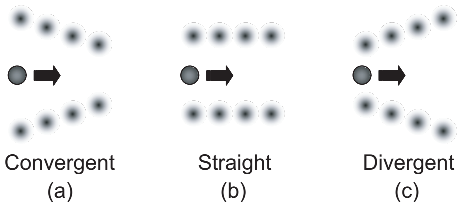

Figure 5. Types of 3D virtual tracks.

Before starting the trials, we oriented subjects to take each decision for pressing a key as perception response with certainty and to perform the selections one after the other trial in a serial fashion.

The evaluation process consists of the following two experimental stages:

- Training stage is an opportunity for subjects familiarizing with the existence of three types of virtual tracks, as shown in Figure 5. The following training tracks are presented to subjects:
  - A converging track deflected by 30^o.
  - A diverging track deflected by -30^o.
  - A straight track (without deflection).
  - All tracks set to the traveling speed of $4\ km/h$.

  In practice, the subject travels along a virtual track and categorizes this track. Subjects select the type of the track by pressing a key on the keyboard as response for a trial. For each trial, the corresponding correct answer is given to the subject.

- In the testing stage, a total of 35 trials for each subject are prepared. A subject travels along the virtual tracks that are set as follows:
  - The natural-sounding tone is adopted as non-speech sound.
  - The traveling speed is set at $4\ km/h$.
  - The values of deflection angle θ shown in Figure 4 can be -30^o, -5^o, -3^o, 0^o, 3^o, 5^o, or 30^o; in other words, we can have seven different virtual tracks.
  - For each deflection angle, the 3D virtual acoustic space system creates five tracks totalizing 35 trials.
  - The sequence of these tracks is randomly arranged to avoid the influence of the learning process by subjects on performance evaluation.

3. First Evaluation Results

Figure 6 shows a comparison of performance results for a typical subject traveling along the virtual tracks in forward and backward direction.

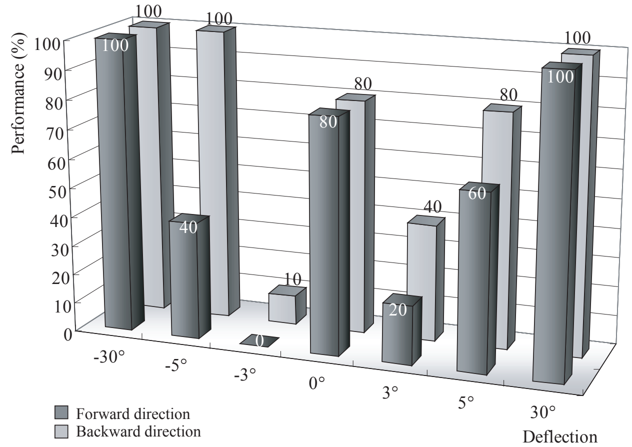

Figure 6. Performance results of a typical subject traveling along the virtual tracks in forward and backward direction.

Table 1 shows the performance results for all subjects traveling at $4\ km/h$ along the tracks in forward and backward direction. For each traveling direction, we have the detailed performance results as follows:

- $Avg1$ represents the average of all results, that is, categorization performance results by all subjects.
- $Std1$ represents the standard deviation in relation to the above calculated $Avg1$ values.
- $Avg2$ represents the average of filtered results considering the values of $Std1$, that is, discarding values out of average deviation from the mean.
- $Std2$ represents the standard deviation in relation to the filtered results represented by $Avg2$.

Figure 7 shows a comparison of $Avg1$ results for subjects traveling in forward direction with results for subjects traveling in backward direction.

Figure 8 shows a comparison of $Avg2$ results for subjects traveling in forward direction with results for subjects traveling in backward direction.

4. Discussion

Table 1 shows that the standard deviation values for backward direction decreased more than the values for forward direction when the performance results were standardized by $Std1$ values. The decrease occurred for all deflections.

In Figure 6, we can verify that the spatial conceptualization performance improved when the subject traveled along the virtual tracks in a backward direction. For example, the performance improved 60% for virtual tracks deflected by -5^o.

Figures 7 and 8 show that subjects' performances improved when the moving direction was reversed from forward to backward. Mainly for delections by -3^o and -5^o

Table 1. Performance results for all subjects traveling along the tracks in forward and backward direction.

	Forward				*Backward*			
Deflection	$Avg_1(\%)$	$Std_1(\%)$	$Avg_2(\%)$	$Std_2(\%)$	$Avg_1(\%)$	$Std_1(\%)$	$Avg_2(\%)$	$Std_2(\%)$
-30^o	93	10	93	10	89	22	94	10
-5^o	36	22	42	18	71	23	69	11
-3^o	26	33	30	12	56	28	45	9
0^o	53	36	57	18	57	25	51	11
3^o	39	30	36	18	46	31	50	11
5^o	61	33	55	18	59	29	60	16
30^o	97	7	97	7	96	12	98	6

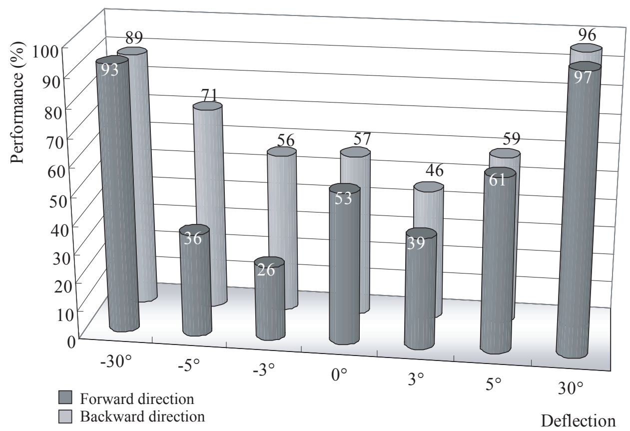

Figure 7. Comparison of $Avg1$ results for subjects traveling in forward and backward direction.

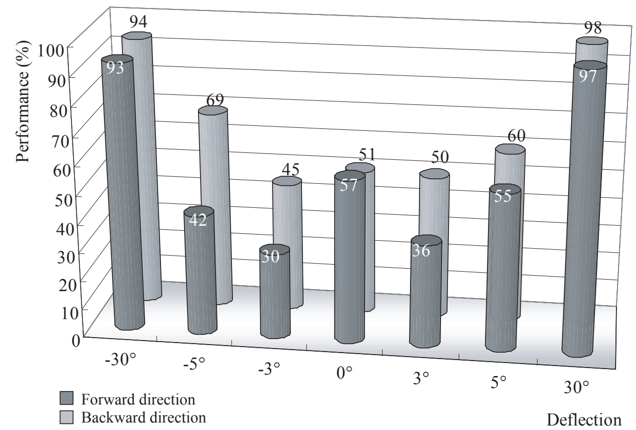

Figure 8. Comparison of $Avg2$ results for subjects traveling in forward and backward direction.

we can verify noticeable performance improvements (respectively, 26 to 56%, and 36 to 71% in Figure 7 and respectively, 30 to 45%, and 42 to 69% in Figure 8).

5. Conclusion

We proposed a new approach using natural-sounding tones as nonspeech sounds for supporting eventual user interfaces to improve their perception of alterations on aural surfaces. By improving their perception ability, the objective was to evaluate the design of an aural interface for enhancing their spatial conceptualization performance.

The evaluation process consisted of subjects hearing the natural-sounding tones and traveling along the virtual acoustic tracks in forward and backward direction.

First evaluation results revealed that the subjects were able to perform their spatial conceptualization tasks for categorizing different types of virtual tracks more effectively and ef□ciently when they traveled in backward direction.

It is surprising that a first evaluation of the new approach for designing aural user interfaces produced so promising results. By considering these results, we concluded that the proposed approach is viable and essential requirement for designing novel aural user interfaces as supporting systems for visually impaired and elderly people.

References

[1] W.W. Gaver, What in the world do we hear? An ecological approach to auditory event perception, *Ecological Psychology*, 5, 1993, 1–29.

[2] W. Buxton, Using our ears: An introduction to the use of nonspeech audio cues, in E. Farrell (Ed.) *Extracting meaning from complex data: processing, display, interaction, Proceedings of the SPIE*, Vol. 1259, 1990, 124–127.

[3] K. Ito, Semantics of TAU, *Gendaishiso Journal* (in Japanese), 1994, 178–187.

[4] H. Donker, P. Klante, and P. Gorny, The design of auditory user interfaces for blind users, *Proceedings of NordiCHI*, Arhus, Denmark, 2002, 149–155.

[5] S. Nomura, M. Tsuchinaga, Y. Nojima, T. Shiose, H. Kawakami, O. Katai, and K. Yamanaka, Novel non-

speech tones for conceptualizing spatial information, *Artificial Life and Robotics* (to be published), 2007.

[6] W.D. Jones, Sight for sore ears, *IEEE Spectrum*, 41(2), 2004, 13–14.

[7] S.D. Erulkar, Comparative aspects of spatial localization of sound, *Physiological Review*, 1952, 1972, 237–260.

[8] N. Fujiki, K.A.J. Riederer, V. Jousmaki, J.P. Makela, and R. Hari, Human cortical representation of virtual auditory space: Differences between sound azimuth and elevation, *European Journal of Neuroscience*, 16, 2002, 2207–2213.

[9] S. Nomura, T. Shiose, H. Kawakami, O. Katai, and K. Yamanaka, A novel "sound visualization" process in virtual 3D space: The human auditory perception analysis by ecological psychology approach, *Proc. of* 8^{th} *Asia Pacific Symposium on Intelligent and Evolutionary Systems*, Cairns, Australia, 2004, 137–149.

[10] T. Shiose, K. Ito, and K. Mamada, The development of virtual 3D acoustic environment for training `perception of crossability', *Proc. of the* 9^{th} *International Conference on Computers Helping People with Special Needs*, Paris, France, 2004, 476–483.

[11] S. Nomura, M. Tsuchinaga, Y. Nojima, T. Shiose, H. Kawakami, O. Katai, and K. Yamanaka, A novel human-environment interface for conceptualizing spatial information in non-speech audio, *Proceedings of the Eleventh Symposium on Artificial Life and Robotics*, Beppu, Oita, Japan, 2006, 305–308.