Visual Resonator: Interface for Interactive Cocktail Party Phenomenon

Junji Watanabe
PRESTO Japan Science and Technology Agency
3-1, Morinosato Wakamiya, Atsugi-shi, Kanagawa,
243-0198, Japan
watanabe@avg.brl.ntt.co.jp

Hideaki Nii
The University of Electro-Communications
1-5-1, Choufugaoka, Choufu-shi, Tokyo,
182-8585, Japan
nii@hi.mce.uec.ac.jp

Yuki Hashimoto
The University of Electro-Communications
1-5-1, Choufugaoka, Choufu-shi, Tokyo,
182-8585, Japan
hashimoto@hi.mce.uec.ac.jp

Masahiko Inami
The University of Electro-Communications
1-5-1, Choufugaoka, Choufu-shi, Tokyo,
182-8585, Japan
inami@computer.org

CHI 2006, April 22–27, 2006, Montreal, Quebec, Canada.
ACM 1-59593-298-4/06/0004

Abstract
We present Visual Resonator (**VR**), an auditory interface that promises to provide an interactive realization of the Cocktail Party Phenomenon. The wearer of this interface can hear a voice or auditory information only from the direction in which he/she is facing, and can send his/her voice only in the direction towards which he/she is facing. In addition, individuals who are within sight of each other can have a conversation, even if they are not close enough to talk directly, since an infrared signal is used to transmit the auditory information.

Keywords
Visual Auditory-interface, Visual Resonator, Direction Selective Infrared Transmission, Cocktail Party Phenomenon

ACM Classification Keywords
H5.1. Information interfaces and presentation (e.g., HCI): Multimedia Information Systems

Introduction
A lot of interfaces have been proposed for the integrative processing of visual and auditory modalities [1-4]. These devices have undoubtedly had to address the fact that the range of information available from each modality is very different. Visual information can be transmitted with varying degrees of selectivity. It

can be sent to a large area simultaneously or to a pin-point by laser beam. On the other hand, auditory information always spreads out in all directions, because it is transmitted through the air without any directionality. If there are many sound sources around the receiver, this feature of auditory information can cause signal/noise problems. For example, conversation has to stop, even in a face-to-face situation, when other people are talking too loudly or traffic is passing. However, a well-known feature of human auditory perception is the Cocktail Party Phenomenon, which enables people to focus their attention on a single individual, typically the person they are talking to. They can pick out the speech of the person they are facing. Encouraged by our knowledge of this phenomenon, we are now proposing an interface called the Visual Resonator, which enhances one's ability to distinguish auditory information from surrounding noise and also allows the wearer's voice to be transmitted solely to the person he/she is facing. The situation realized by this interface can be thought of as an interactive Cocktail Party Phenomenon.

Visual Resonator

VR is an integrated unit consisting of a microphone, headphones, infrared transmitter and receiver as shown in fig. 1. When the interface receives an infrared signal, the received signal is sent to the headphones and translated into sound. An important characteristic of this interface is that the infrared receiver is fixed to the head and has directional selectivity. Therefore, signals can only be received from the direction in which the wearer is facing and he/she can only hear a voice or other auditory information that originates from that direction. This interface can also be used to transmit speech that can be selectively received by other people. The microphone records the voice and translates it into an infrared signal which is then transmitted like a laser beam in the direction that they are facing.

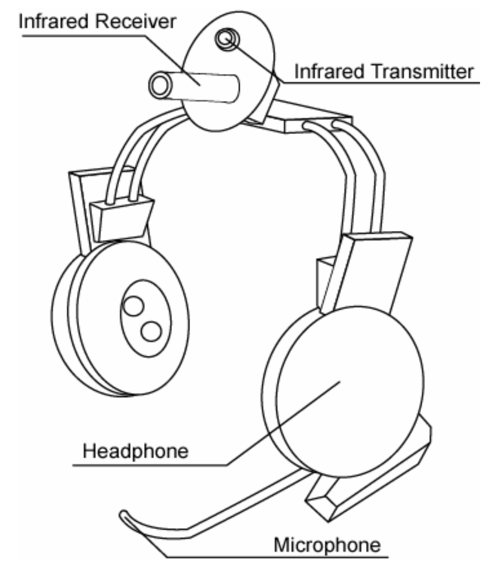

figure 1. This interface is an integrated unit composed of a microphone, headphones, infrared transmitter and receiver.

The transmitted signals can only be received by the person who is facing the speaker as in fig. 2. Thus, people who are within sight of each other can have a conversation, even though they are not close enough to talk directly. In addition, the wearer could obtain recorded auditory information from an object that he/she is looking at as in fig. 3, since the infrared transmitter can be attached to an object, instead of to a person. By combining these two approaches, speech that has been transmitted by an individual in the direction he/she is facing could be relayed by an infrared transmission unit at the object and then received by other people, who are also viewing the same object. Thus, people looking at the same object could communicate with each other through this interface as in fig. 4, as if the voice signal were reflected off the object. An interface of this kind would facilitate communication between people who are looking at the same object and have shared interests.

The directional selectivity of the infrared signals referred to above coincides with the head orientation of the wearer. Gaze is generally a good indicator of a person's attention to external objects. When humans pay attention to an external object, they usually make eye movements towards the object of their interest so as to have it in the center of their visual field. Although the eyes are the primary source for determining a person's focus of attention during social interaction [5,6] and many gaze-based interfaces have been realized [7], the focus of attention also depends on the direction of his/her head, body posture and other gestures such as pointing [8]. In fact, head orientation has also been used for defining focus of attention [9,10], and for a PC's joystick [11,12]. In addition, it is preferable that the interface is simple and easy to wear. Therefore, we have used the wearer's head orientation as an indicator of his/her focus of attention.

Implementation

This interface was implemented by modifying a commercial wireless headphone system (Audio-technica ATH-CL33). The existing infrared sensor was replaced with an infrared receiver surrounded by a long metal cylinder in order to achieve directional selectivity as in fig.1. The length of the metal cylinder was either 9 or 12 cm for 60 or 40 degrees of receivable angle, respectively. The existing transmitter was modified to send infrared signals over a narrow range, the infrared signal being transmitted by Frequency Modulation. So, even if a number of auditory signals were received at the same time, the receiver would concentrate exclusively on one auditory signal.

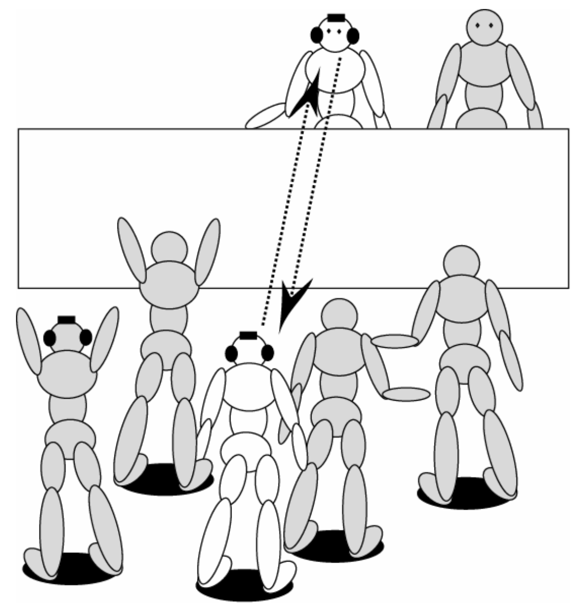

figure 2. Example of interactive Cocktail Party situation.

Scenario 1: Interactive Cocktail Party situation (figure 2)

Junji is attending a dance party. Several **VRs** are delivered to the attendees including Junji. It is very crowded on the dance floor and techno music is playing at full blast. He can't move or make conversation. He notices that the DJ is an old friend. They catch one another's eye and turn their heads towards each other. Junji speaks to the DJ "Hello! Is that you John?" DJ, John answers "Yes, Junji!" They are able to talk, even in a loud music environment, because both are wearing **VR** and **VR** uses infrared signals to transmit auditory information. Then Junji requests his favorite song through **VR**.

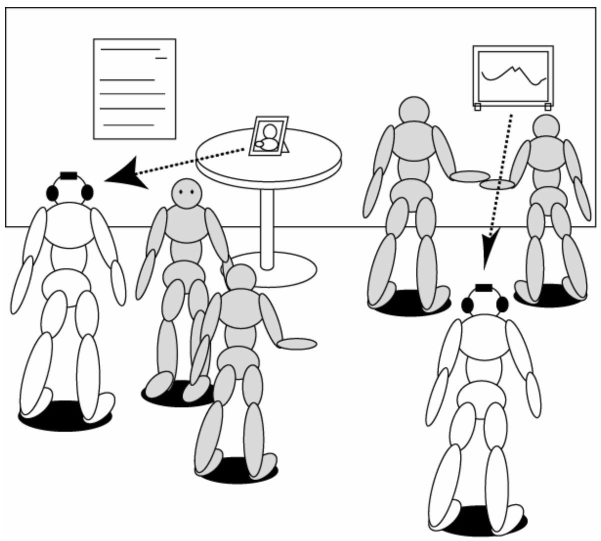

figure 3. Example of head-oriented information presentation.

Scenario 2: Head-oriented information presentation (figure 3)

Junji visits an art gallery and borrows **VR** at the entrance. A lot of visitors are crowded round a famous painting. Junji cannot get a close look, but when he looks at it from a distance, he is able to hear a commentary about the work of art. An infrared transmitter has been attached to the work of art and **VR** receives the signal, when Junji looks at it. Although various information presentation devices have been introduced in museum exhibitions [3], our interface would allow visitors to access a commentary about an exhibit merely by facing it, rather than by pushing buttons. Wearers would automatically hear auditory information from whatever they were looking at and whatever their focus of interest was. In addition, the sound source could be attached to any kind of object, such as clothes or even moving objects.

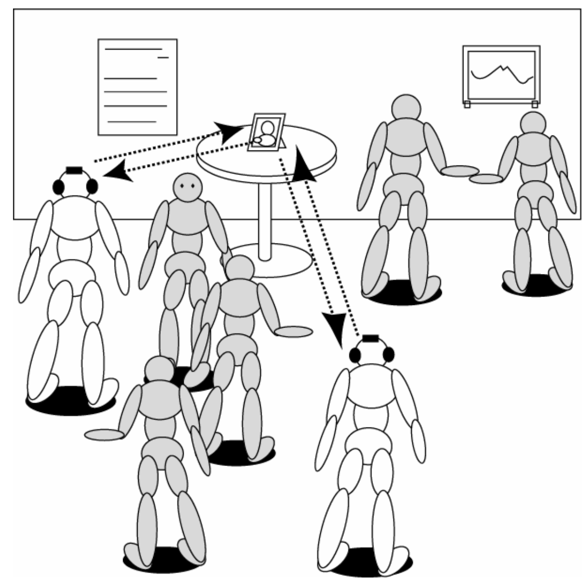

figure 4. Example of conversation through shared interests.

Scenario3: Conversation through shared interests (figure 4)

Junji is pottering around a gallery, when he sees what he considers to be one of the best works and suddenly picks up another person's voice, "Wow, this is great!" The voice has come from another visitor, who is also wearing **VR**. This visitor is standing on the other side of the gallery. Through **VR**, the two of them are able to exchange opinions quietly (without disturbing other visitors). Then, they are able to have a direct conversation about the artwork once they are outside the gallery. The recorded voice was relayed by the infrared transmission unit at the artwork, and then was received by other people, who are looking at it. Thus, this interface can promote communication between people with the same interests.

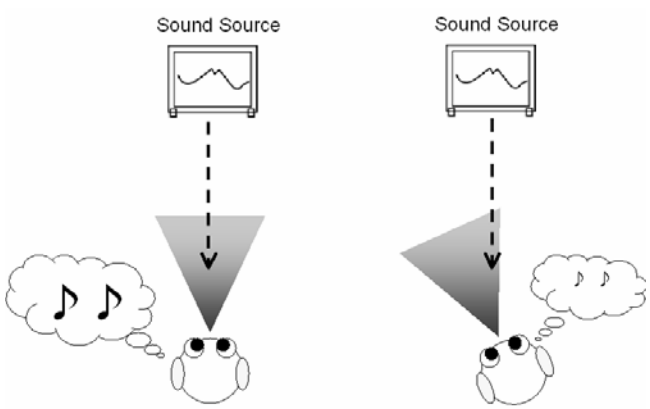

figure 5. Volume change according to the relative angle between head orientation and sound source

Future Works

Using **VR** as implemented so far, the wearer will always hear auditory information at a fixed volume, when VR receives the infrared signal. However, if the volume changed according to the relative angle between the orientation of the head and the sound source as shown in fig. 5, the wearer could readily identify where the auditory information comes from. In future, we intend to introduce volume changes which will help the wearer to localize the sound source.

In addition, as an application of **VR**, art work on the visual and auditory communication can be made [13]. This kind of technology-based art works can reveal new possibilities of media arts.

Discussion

Recently, auditory information presentation methods using visible or invisible light have been proposed. Traffic lights using LEDs can also be used for presenting other kinds of information in addition to traffic signals. CoBIT, proposed by Nishimura et al. [3], uses weak infrared signals and can work without batteries. Although the range over which the signal can be heard is limited, it can be an economical system. Sound Scope Headphone by Hamanaka [4] is one of auditory scene retrieval. The wearer can compose music by rotating his/her head. However, neither of these systems has an interactive aspect; they only hear the sound. An important characteristic of **VR** is that the wearers have interactive, directional selectivity, and therefore the wearers can only hear sound from the direction they are facing and can only send their voice towards the area they are looking at. They can have an interactive visual-auditory experience.

Conclusion

In this paper, we presented Visual Resonator (**VR**), an auditory interface that realizes an interactive version of the Cocktail Party phenomenon and promotes communication not only between people who are within sight of each other but also between people with a common interest.

Acknowledgements

This research is supported by PRESTO, Japan Science and Technology Agency.

References

[1] Basu, S., and Pentland, A. Smart headphones. *Ext. Abstracts CHI2001*, ACM Press (2001), 267-268.

[2] Smith, D., Donald, M., Chen, D., Cheng, D., Sohn, C., Mamuji, A., Holman, D. and Vertegaal, R. OverHear: augmenting attention in remote social gatherings through computer-mediated hearing. *Ext. Abstracts CHI2005*, ACM Press (2005), 1801-1804.

[3] Nishimura, T., Itoh, H., Nakamura, Y. and Nakashima, H. A compact battery-less information terminal for Interactive information support, *The Fifth Annual Conference on Ubiquitous Computing 2003*, Workshop: Multi-Device Interfaces for Ubiquitous Peripheral Interaction (2003).

[4] Hamanaka, M.

http://staff.aist.go.jp/m.hamanaka/index.htm (Feb. 10th , 2006)

[5] Nielsen, J. Noncommand user interface. *Communications of the ACM*, 36, 4 (1993), 83-99.

[6] Vertegaal, R., and Ding, Y. Explaining effects of eye gaze on mediated group conversations, In *Proc. CSCW 2002,* ACM Press (2002), 41-48

[7] Jacob, R. The use of eye movements in human–computer interaction techniques: what you look at is what you get, ACM Transactions on Information Systems 9, 3(1991), 152–169.

[8] Langton, S.R., Watt, R.J., and Bruce., V. Do the eyes have it? cues to the direction of social attention. Trends in Cognitive Science, 4, 2 (2000) 50-59.

[9] Stiefelhagen, R., Yang, J. and Waibel, A. Estimating focus of attention based on gaze and sound In *Proc. PUI 2001*, ACM Press (2001), 1-9

[10] Skaburskis, A.W., Shell, J.S., Vertegaal, R., and Dickie, C. AuraMirror: artistically visualizing attention. *Ext. Abstracts CHI2003*, ACM Press (2003), 946-947

[11] Lin, M., Radwin, R., and Vanderheiden, G. Gain effects on performance using a head-controlled computer input device. *Ergonomics*. 35, 3(1992), 159-175.

[12] Evans, D.G., Drew, R., and Blenkhorn, P. Controlling Mouse Pointer Position Using an Infrared Head-Operated Joystick *IEEE Transaction on Rehabilitation Engineering*, 8, 1 (2000), 107-117.

[13] Watanabe, J., and Sugimoto, M. Visual Resonator, Ars Electronica Cyber Art 2004, NextIdea (2004), 301.