resonEars: Amplifying perceived audio information by gaze shifting

Jifei Ou
ROOM 348P, 75 AMHERST
ST, CAMBRIDGE, MA 02139
jifei@media.mit.edu

ABSTRACT
In this paper, I present an interaction mode that amplifies different sound by gazing at its source. Human ears are not good at focusing/defocusing on a certain sound. This causes distraction. I envision a system that uses gaze as guide to shift amplified sound in the physical world. I also demonstrate an working art installation that allows people to experience this interaction mode.

Author Keywords
Gaze, auditory augmented reality, attention.

INTRODUCTION
Unlike eyes, human ears constantly perceive information from every direction to the body. Our ears are good at distinguishing the location of the sound source, but very limited at focusing on one sound from background noise. With eyes, we can easily focus or defocus on a object in the environment, or move head around to maximize our attention. However we cannot completely shut down or actively tune our auditory sense.

In perception theory, the cocktail party effect[1].describes the ability in perception to select one desired sound from a background of ambient noise. In a crowded situation, many voices are speaking simultaneously, but we can 'focus' our ears on one voice and ignore others that are equally strong. However this needs a lot of effort and people can still easily be distracted by other sound.

In this paper, I envision a system that can amplify a desired sound by using gaze as guide. An artistic installation that allows people to experience look-to-listen interaction mode is presented as well.

RELATED WORKS
Sound as one crucial topic has been largely explored in the field of VR. Most of researches focus on how sound helps us better establishing a sense of space and orientation. Based on those researches, many applications have been also proposed to facilitate a new mode of human-computer interaction (HCI).

Taking advantage of human abilities of simultaneous listening and memory of spatial location, Dynamic Soundscape[2] is a browsing system that provides a spatial interface for temporal navigation of audio data. users navigate through one long audio file by turning a knob. Audio Hallway[3] is another project that helps user to browse collections of related audio files. User travels up and down the Hallway by head motion, passing "rooms" on the left and right sides. In the field of HCI, gaze as an interaction cue has been largely discussed as well. In the paper[4], MIT AI lab evaluated how look-to-talk paradigm is indeed a natural alternative to speech.

USER SCENARIOS
Based on previous research, I propose using gaze as a guide to amplify a desired sound in physical environment. This interaction mode would benefit various scenarios and situations:

Live Concert

A live concert is very much enjoyable to most of people. However it is difficult for common people to extract one single instrument from the whole music piece. In this scenario, user gazes on a certain musician. What he plays will be amplified and transmitted to the users earphone. This scenario can be not only a new experience design, but also useful for music education.

Figure 1

Browsing Large Amount of Video Data

Video is a medium that mixes spatial and temporal. Browsing large amount of video data can be time consuming. In this scenario, user can play multiply tracks of video on different screens. When he gazes at one screen, the corresponded audio will be amplified so that he can get the details of the video.

Figure 2

Augmented Painting

In the world of art, paintings are usually considered as static and quiet. Can we augment painting with sound to transform it into an immersive storytelling tool? By adding the system, visitor will hear a soundtrack spreading on different locations of a painting. When he focuses on one location, the corresponded audio will be amplified.

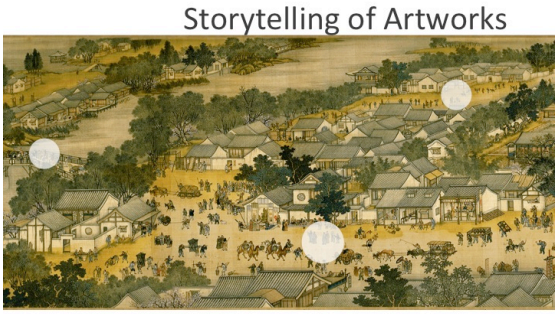

Figure 3

Observing in Nature

Can we also actively tuning the sound of nature? In this scenario, the system is implemented in a natural environment. When people walk through, he will hear all kinds of sound from birds, insects, wind, river, etc.. By looking at them, the sound of bird or other source will be amplified and transmitted to the user.

IMPLEMENTATION

I implemented an artistic installation to let people experience the interaction of "look-to-focus" on audio information. The system consists 8 speakers and 4 computers to playback and alternate sound. A Microsoft Kinect is used for head position tracking to approximately archive gaze tracking.

Visitor stands in front of Kinect and hear an acapella choir song coming from 8 speakers. Each speaker plays one voice. When visitor move his head towards one speaker, its sound volume will increase so that visitor can hear more from this voice without completely diminish others.

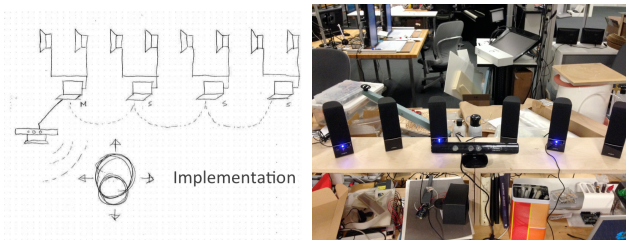

Figure 4

FUTURE WORKS

The art installation provides limited experience. In the future, I will try to implement a functional system that allows user move freely. There are three aspects that need to be implemented

1. Sensing. Embedding microphones in physical environment in a specific situation, for example, live concert.
2. Augmentation. Using transducer earphone for user to receive the amplified sound. The transducer earphone acts as an "overlaid glass" of ears, so that user can still hear the sound from environment.
3. Gaze detection. Using IR transmitter and receiver to detect if user is looking at a certain object.

REFERENCES

1. Handel, S. Listening: An Introduction to the Perception of Auditory Events. MIT Press, 1989
2. Kobayashi. M; Schmandt C., Dynamic Soundscape, Mapping Time to Space for Audio Browsing. In Proc. CHI 1997, ACM Press (1997), 194-201
3. Schmandt C., Audio Hallway: a Virtual Acoustic Environment for Browsing. In Proc. UIST 1998, ACM Press (1998), 163-170
4. Oh A.; Fox H.; Van Kleek M.; Adler A.; Gajos K.; Morency L.; Darrell T., Evaluating Look-to-Talk: a Gaze-Aware Interface in a Collaborative Environment In *Proc. CHI EA 2002*, ACM Press (2002), 650-651.