Proposal and Evaluation of Mapping Hypermedia to 3D Sound Space

Radoslav Buranský

Faculty of Mathematics, Physics and Informatics
Comenuis University, Bratislava

Abstract

In this paper we present a new way of representing information contained in HTML pages. The goal is to map data to 3D sound space and evaluate this mapping. We were mostly inspired by Goose and Möller [1], because we think that they have found practical solutions for many problems in non-visual browsing. What follows is a brief overview of methods we are going to use. After implementation they will be evaluated by blind users.

Introduction

Nowadays most web browsers for the blind use only one channel audio output. This is very constraining, because a lot of additional information can be provided by positioning of sound in space. For instance we use azimuth (position in horizontal plane) to indicate position in document. If sound comes from the left, then the read text is on the beginning of document, if it comes from the right, it is at the end. Another good reason is that position in space helps human's memory to move information from short-term to long-term memory [2]. It helps organizing new learned knowledge into hierarchical structure.

Usage of overlapping sounds is challenging. Human's ability to listen to one from many concurrent sounds is known as the "cocktail party effect" [4]. But the sounds must differ enough in some parameters like volume, pitch or position in space. We will try to use this effect to make browsing faster.

We want to create a real working solution, so it doesn't have high hardware requirements. Even a multi-channel audio system is not necessary, because 3D sound can be simulated using a common headphones [3]. We use some open source projects. They provide functionality such as 3D audio output, speech synthesizing or HTML parsing.

The browser is fully controlled by a keyboard. Recommended output device are headphones plugged into a sound card with HRTF functionality.

Browsing

Let's start talking about the way we represent data using audio space. We decided to place sounds to horizontal plane only and most of them are in the front half-plane. This is because there are many front-back and top-down confusions reported in previous researches by users [5,6]. Front-back confusion means, that user can not correctly tell whether the sound comes from in front or behind him. Top-down confusion is similar. So sound sources are placed at reduced semi-circle, called Stage-Arc. (fig. 2). The reduction is caused by lower accuracy of human's ear to azimuth change in both extreme positions (fig. 1). It can be seen in figure 2 that the Stage-Arc does not cover whole 180°. It is reduced by 20° at both ends.

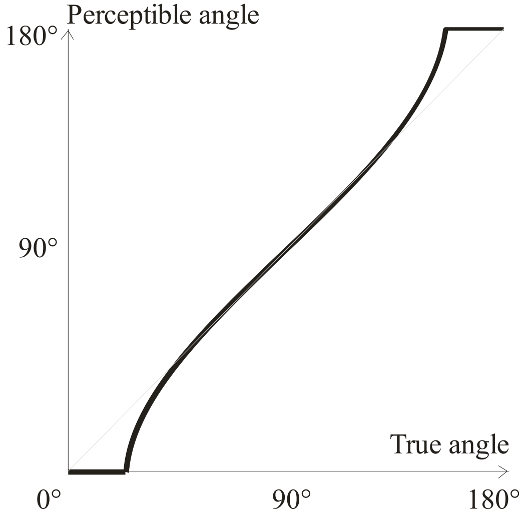

Figure 1: The curve perceptible to human.

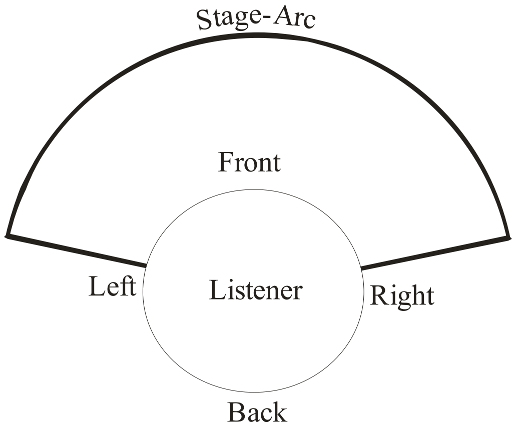

Figure 2: Eliminating the imperceptible regions of the arc.

We use four different synthesized male voices (speakers). First reads headings, second reads contents, third is for links and the last one belongs to images. Reading images means to read their alternative description. It is up to web developer to provide such information. All voices are located at the Stage-Arc right in front of listener. They don't move, because we think that it is not pleasant when a text is read from one side for a long time. The reason why to use more voices is that human can easily recognize which voice belongs to which speaker. It is sufficient just to read a text and user immediately knows if it is a heading or a link. Actually we use Festival open source project to synthesize speech. We are restricted to use four male voices as they are provided with Festival. To create more voices is out of scope of our project.

Along with synthesized speech we use non-speech sounds called earcons. An earcon is any sound that does not contain human voice neither synthesized nor prerecorded. It is a sound symbol with exact meaning. Four earcons are used with the same meaning as voices. So we have one earcon for headings, one for contents, another for images and the last is for links. When a text is being read then earcon is sounded from the relative position along the Stage-Arc. It provides information about position in document. Its position on the Stage-Arc is relative to position in document.

Imagine a situation that a heading located at the beginning of document is being read. Then the voice for headings reads this heading and in parallel the heading earcon is sounded. The voice is located in the center, but the earcon is on the leftmost side of the Stage-Arc telling user that the heading is at the beginning. The same holds for images, texts and links.

Although every visual browser views a document in 2D plane, any hypermedia document can be linearized. Imagine that this linearized document is simply stretched to the Stage-Arc. The beginning is on the left and the end is on the right side. That is the way our browser works. Every object (heading, link, image or a text) has its exact position in the document and so it has exact position on the Stage-Arc. For example a heading located at 20% from the beginning of a document is placed at 20% of the Stage-Arc (0% is on the leftmost side).

One of basic problems in hypermedia systems for visually impaired is the way how links are represented and followed. Links are of two types: intra-document and inter-document. If these two links are not represented in different ways, user can get lost in hyperspace. The way we solved this problem follows. Intra-document link uses three sounds: take off, flying and landing sound. When user activates this type of link the take off sound is played from the relative position on the Stage-Arc. Then flying sound moves towards target and at last landing sound is played (fig. 3).

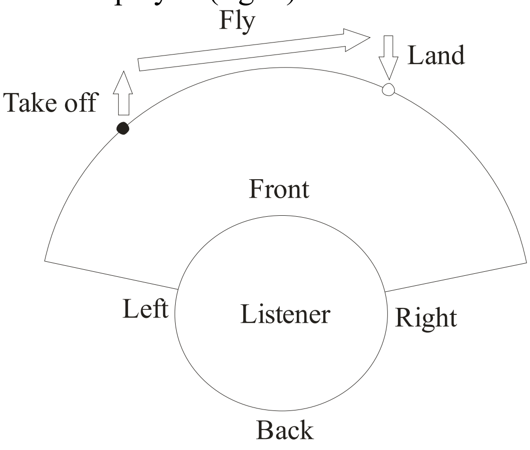

Figure 3: 3D audio intra-document link traversal.

For inter-document link we use only two sounds: take off and landing. These are different from sounds used in intra-document link. Take off sound fades out and its distance from listener grows. Landing sound comes nearer and becomes louder (fig. 4).

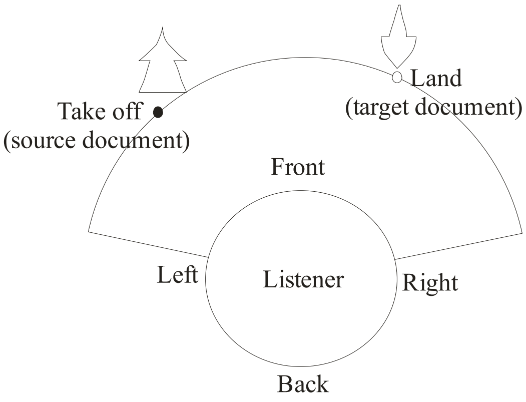

Figure 4: 3D audio inter-document link traversal source.

Another problem is activating a link. We mark link as active when it is being read. To mark link as active is only internal action of the browser. No special sound is played. It remains active until another link is encountered in the document. User can follow active link at any time by pressing space bar on keyboard.

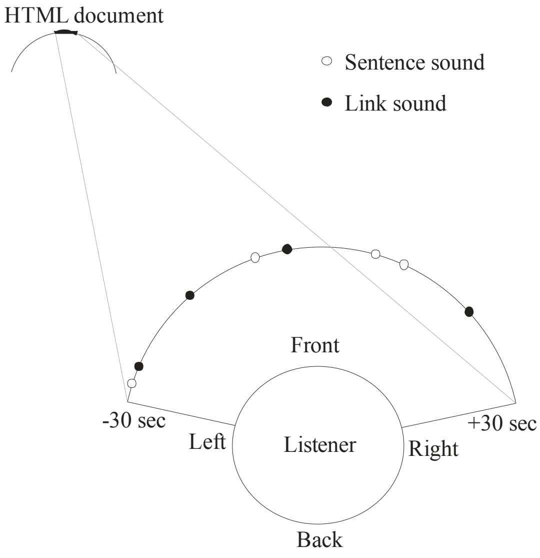

Figure 5: 3D sound survey of a region of the document.

When sighted users browse, they firstly create quick overview of whole document by looking at it. This is what blind users cannot do. We provide functionality by creating a sound survey of the area surrounding the user's current position in the document. The user can request this survey at any time.

The user can specify in time units how far in both directions the sound survey should extend. From figure 5 it can be seen that the surrounding region (in this example ± 30 seconds) is magnified and stretched to whole Stage-Arc. The survey is faster than real time (60 seconds long area can be heard in 15 seconds). No text is read during survey, only the four earcons are used.

We provide also global sound survey. This takes advantage of "cocktail party effect" and the knowledge of perceptual importance. The same four voices (readers) are used, but they are located in the front, at the right side, behind, and at the left side from the listener. Only headings of the document are being read. For the global survey, all four voices read headings. The first one is read by the first reader located in the front of the listener. After few seconds the sounds starts to fade out and the next reader, on the right, is activated. He reads next heading and the sound is louder then of the previous one. After few seconds the same action happens again, and the third speaker is activated. At this time three speakers are active, but the last one is dominant. We think that in this way, user can get the fastest overview of whole document. In fact, user gets all headings of a document in a few seconds. First tests show that this idea is helpful.

Global navigation has four modes. In the first mode whole document is read. The second mode reads only links. The third reads only headings and the fourth reads only textual content without other structural elements. User uses keyboard to invoke actions. Arrow buttons are used to move to previous/next object in the document. “Home” and “end” buttons are used to jump to the beginning or to the end of a document. “Page up” and “page down” buttons skip some objects. Space bar activates the active link (if any) and starts navigation to the link’s destination document.

Conclusions

We have described design of non-visual interactive audio browser which should help blind users. All these ideas will be implemented and evaluated in next few months. After comments and suggestions taken from visually impaired users, we will improve our model.

References

[1] Goose, S. and Möller, C., A 3D Audio Only Interactive Web browser: Using Spatialization to Convey Hypermedia Document Structure., ACM

[2] Kobayashi, M. and Schmandt, Ch., Dynamic Soundscape: mapping time to space for audio browsing, ACM, 1997

[3] Gardner, B. and Martin, K., HRTF Measurements of a KEMAR Dummy-Head Microphone, MIT Technical Report #280, 1994

[4] Schmandt, Ch., and Mullins, A., AudioStreamer: Exploiting Simultaneity for Listening, ACM, 1995

[5] Wenzel, E. M., Wightman, F. L. and Kistler, D. J., Localization with non-individualized virtual acoustic display cues, ACM, 1991

[6] Lokki, T., Gröhn, M., Savioja, L. and Takala, T., A Case Study of Auditory Navigation in Virtual

Acoustic Environments, Helsinki University of Technology