An Audio-Haptic Interface Based on Auditory Depth Cues

Delphine Devallez
University of Verona
Strada Le Grazie 15
37134 Verona, Italy
delphine.devallez@univr.it

Federico Fontana
University of Verona
Strada Le Grazie 15
37134 Verona, Italy
federico.fontana@univr.it

Davide Rocchesso
IUAV of Venice
Dorsoduro 2206
30123 Venezia, Italy
roc@iuav.it

ABSTRACT

Spatialization of sound sources in depth allows a hierarchical display of multiple audio streams and therefore may be an efficient tool for developing novel auditory interfaces. In this paper we present an audio-haptic interface for audio browsing based on rendering distance cues for ordering sound sources in depth. The haptic interface includes a linear position tactile sensor made by conductive material. The touch position on the ribbon is mapped onto the listening position on a rectangular virtual membrane, modeled by a bidimensional Digital Waveguide Mesh and providing distance cues of four equally spaced sound sources. Furthermore a knob of a MIDI controller controls the position of the mesh along the playlist, which allows to browse the whole set of files. Subjects involved in a user study found the interface intuitive and entertaining. In particular the interaction with the stripe was highly appreciated.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Experimentation, Human Factors, Performance

1. INTRODUCTION

1.1 The Use of Distance Information

While most research dedicated to the design of new auditory interfaces focuses on directional spatialization of multiple sound sources [12, 16, 13, 1, 2], we propose an interface based on distance information. The motivation is driven by the ability of depth information to provide a hierarchical relationship between objects and therefore bring the attention of the user on the closest sound source while still hearing the other ones in the background. We believe that auditory interfaces would take great advantage of the depth dimension to provide additional information on the spatial layout of sound sources and therefore improve the organization of the auditory scene. In 1990, Ludwig [7] already suggested that techniques used in the music industry, such as reverberation and echo, could be valuable to the ordering of multiple sound sources in auditory interfaces.

Auditory distance cues include intensity, direct-to-reverberant energy ratio, spectrum and binaural cues (Refer to [18, 10] for detailed reviews of distance cues). Their respective contributions to distance perception may differ according to the nature of and the familiarity with the sound source and the environment, as well as the availability of other non-acoustical cues. In general, distance perception is much less accurate than directional localization. Nevertheless, we are rather interested in the perception of the relative distances between multiple sound sources, assuming that the user is able to discriminate between the respective positions of the sound sources. The studies of Strybel and Perrott [15] and Zahorik [19] touch upon the human perception of distance changes by measuring the resolution of source distance with the intensity cue and the direct-to-reverberant energy ratio cue respectively. Both studies suggested the possibility that the aforementioned cues represent distance changes, although the threshold of the direct-to-reverberant energy ratio is much coarser than the threshold of the intensity cue.

1.2 Related Work

With the ability to manipulate the spatial relationships between sound sources in depth, users may be able to focus their attention on a specific sound source corresponding to the closest distance to him or her. This manipulation is closely related to the technique called *acoustic zooming*, which accomplishes the focus on a specific object out of a multitude of sound sources. This is generally done by increasing the intensity of an object, for example as a function of its relative distance to the pointer controlled by the user [17, 11], or as a function of the direction of the source compared to the user's median plane [13, 2]. The *Sonic Browser* developed by Fernström and Brazil [2] is a representative application of the use of acoustic zooming for audio browsing. The spatial location of sound sources is rendered by stereo panning, i.e. loudness difference between the left and right channel. Further assistance to sound localization is given by a function that defines the range of perception, called *aura*. More precisely, it allows to get a finer discrimination between adjacent sound sources, acting like a zoom, in addition the radius of the aura is controllable by the user. The evaluation of the prototype [3] showed that simultane-

ICMI'08, October 20–22, 2008, Chania, Crete, Greece.

ous rendering of multiple stereo-spatialized sound sources allowed to complete the task faster than single-stream rendering. Without any spatialization layer it was shown that multiple concurrently presented earcons jeopardize the identification of the menu items [8]. On the contrary, using the human ability to localize sound sources in space may offer benefits to multi-stream rendering for identification tasks. In their study, Pitt and Edwards [11] showed that a linear menu with 2 to 8 sounds presented simultaneously and panned between the left and right channel produced a fastest average search time compared to the condition where only the sound associated to the pointer's position was played at a time. They also suggested that the task was more natural with concurrently presented sound sources, similar to identification of sound sources in real life situations.
Instead of manipulating directly distance cues such as the intensity, which may also be manipulated by the user and therefore make this cue unreliable, spatialization techniques may offer better alternatives for distance rendering by making the dynamic auditory environment more natural. In particular, recent studies have shown the ability of physical modeling of acoustic propagation to simulate acoustical environments. Of interest is the study conducted by Fontana and Rocchesso [5] which has underlined the effectiveness of a Digital Waveguide Mesh (DWM) modeling a rectangular parallelepiped to provide acoustic depth cues. Listening experiments using this model have shown its ability to render the apparent distance of sound sources. The only drawback of the proposed model was the high amount of computational resources required to simulate the environment, which did not allow a realtime application.

2. THE USER HAPTIC INTERFACE

2.1 Design Approach

This paper presents an interface meant to navigate among multiple audio streams ordered in depth and using a haptic input from the user. Auditory depth information is provided by a Digital Waveguide Mesh modelizing a bidimensional membrane and is presented in section 3. The user haptic interface consists of a ribbon controller, the Infusion Systems *SlideLong*, inspired by music controllers.The choice of a linear position touch sensor is validated by its common use for browsing through the playlist on portable music players equipped with touch sensors. In addition, the *SlideLong* has an active area of $210 \times 20\,\mathrm{mm}^2$, and therefore exhibits a main dimension along which distance information may be explicitly understood. Since the ribbon has a rectangular geometry, it also allows an easy analogy with the geometry of the rectangular DWM simulating the auditory environment. The controller gives a value depending on the position at which the touch is made. Therefore, the continuous interaction is performed by a mapping of the position of the user's finger on the ribbon onto the position of a virtual microphone on the membrane.

2.2 Prototype Development for Audio Browsing

The position tactile sensor gives a value corresponding to the touch position on the ribbon. A gamepad plays the role of sensor interface and is connected to the USB-port of the computer. As underlined by Jensenius et. al [6], such a game controller has the advantages of being cheap

Figure 1: *Picture of the setup.*

and having analog inputs which comply to the 0-5 volt sensor outputs. Besides connecting the sensor outputs on the motherboard is easy and the device uses the Human Interface Device driver supported in Max/MSP[1], which allows to make a fast, simple and low-cost interface sensor out of the game controller. The incoming values in Max/MSP are read by the built-in *human interface* object and are scaled to float numbers between 0 and 255. Therefore, after rescaling, the incoming value from the ribbon controller provides the listening position input to the computation of the auditory signals in the DWM. In this way, a coherent mapping is performed between the touch position on the ribbon and the position of a virtual microphone in the DWM, and by moving the finger on the ribbon the user may explore the virtual environment where different audio streams are being attributed different positions. Like music controllers, this touch sensor intends to provide an interface that is intuitive to use with immediate and coherent response to user's gesture.
Due to restrictions of hearing discrimination capability, it was found by informal listening tests that at most four audio streams could be reproduced in the virtual environment. Therefore a second haptic interface is added in order to allow to browse among more sound files. For this purpose, we use one of the rotary encoders of a keyless MIDI controller (the Novation *ReMOTE ZeRO SL*), which is available at our laboratory. The controller is connected to the USB-port of the computer. The encoder has discrete steps and therefore may be easily manipulated to switch between discrete levels. Like the position tactile sensor, its output value is read in Max/MSP, and determines the four audio streams processed in the virtual environment. Figure 1 shows the complete setup.

3. MODELING OF THE AUDIO SPACE

3.1 The Acoustic Environment

Our proposed virtual acoustic environment consists of a rectilinear two-dimensional mesh whose digital waveguides simulate acoustic wave transmission between each internal

[1] `http://www.cycling74.com/products/maxmsp`

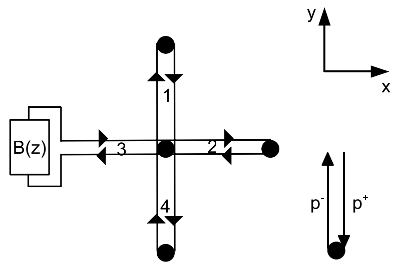


Figure 2: *Zoom on a node which is connected to other odes via waveguides 1, 2 and 4. Waveguide 3 leads to a partially absorbing boundary. Triangles filled in black represent oriented unit delays.*

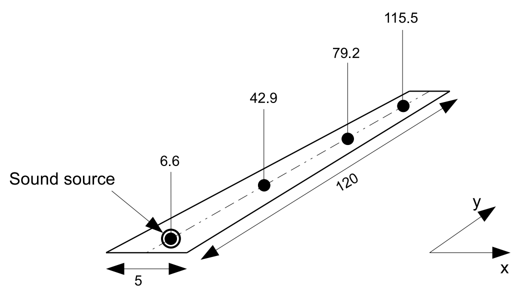


Figure 3: *The virtual membrane with the sound source and the measurement positions. All sizes are in centimeters.*

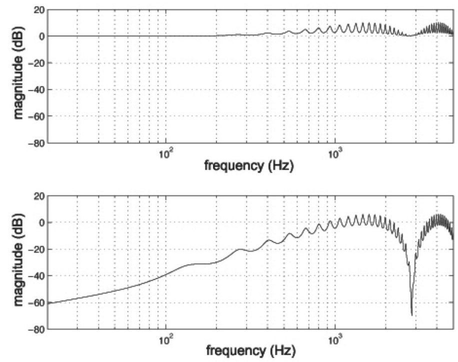


Figure 4: *Frequency responses up to 5 kHz on the membrane . Top: 6.6 cm. Bottom: 115.5 cm.*

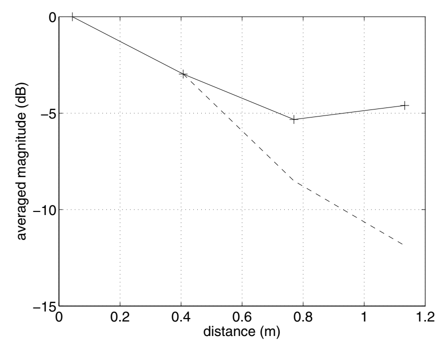


Figure 5: *Average magnitude of the impulse response as a function of distance. Solid line: 2D mesh. Dashed line: Reference open space.*

junction. Each waveguide models the wave decomposition of a pressure signal p into its wave components p^+ and p^-, and each lossless node scatters 4 input signals coming from orthogonal directions, p_1^+, ..., p_4^+ into corresponding output signals p_1^-, ..., p_4^- (see Fig. 2). The properties of the wall materials contribute to the acoustics of a 3D space. This is also the case for a bidimensional acoustic environment since horizontal waves interact with the surface boundaries. Reflections from the boundaries are modeled by Digital Waveguide Filters (DWF), whose coefficients have been tuned to model specific reflective properties of real surfaces [4]. Finally, the number of nodes can be converted into the corresponding membrane dimensions once the speed of sound and the sampling frequency of the simulation have been determined. Due to software restrictions, the mesh dimensions are chosen to be 110×5 nodes, which correspond to a $120 \times 5\,\mathrm{cm}^2$ membrane. The model has been implemented in Max/MSP as an external object for realtime simulations.

3.2 Acoustical Properties of the Membrane

In order to investigate the auditory distance cues inside the virtual environment, impulse responses are computed at different distances on the membrane, corresponding to the positions of the four audio streams. The sound source is assumed to be point-wise and is located on the y-axis of the mesh, at 6.6 cm from the boundary. Measurements of impulse responses are carried out at four positions, on the y-axis of the membrane. Refer to Fig. 3 for the source and measurement positions. Simulation of the listening environment was carried out in `Matlab`. Figure 4 shows the frequency responses up to 5 kHz, measured respectively at 6.6 cm and 115.5 cm on the virtual membrane .

3.2.1 Overall Intensity Level

Figure 5 shows the variation of the total energy with distance. By comparing with the energy decrease in open space, characterized by the well-known 6 dB law, it can be seen that the overall intensity on the membrane decreases significantly less. This behavior allows to hear even the farthest sound sources at any location on the membrane. In addition, the volume can be manipulated by users, which might make the intensity cue unreliable for judging distance. Another reason for limitating the intensity cue for distance judgment is that the level of direct sound varies both with distance and with the energy emitted from the sound source, so that the listener needs some a priori knowledge about the sound source level in order to evaluate its egocentric distance.

3.2.2 Direct-to-Reverberant Energy Ratio

The virtual acoustic space we have designed aims at rendering depth information mainly thanks to the direct-to-reverberant energy ratio cue. For each impulse response, the delay of the direct sound is deduced from the distance between the sound source and the listening point, and is removed from the impulse response. Afterwards the direct energy is integrated among the first 2.5 ms of the delay-free impulse responses, which approximates the duration of Head Related Impulse Responses measured in anechoic conditions and therefore captures the direct path of the

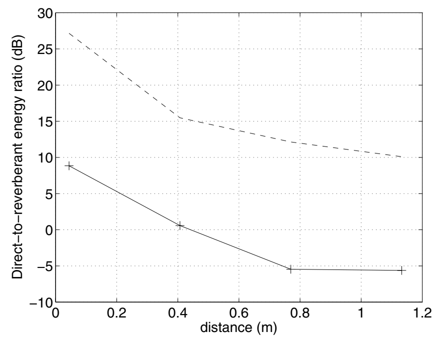


Figure 6: *Direct-to-reverberant energy ratio as a function of distance. Solid Line: 2D mesh. Dashed line: Approximation of a natural environment.*

sound signal [9]. Finally the reverberant energy is calculated from the tail of the delay-free impulse responses. Figure 6 shows the values of the direct-to-reverberant energy ratios for different distances on the mesh. For comparison the direct-to-reverberant energy ratio v was computed for a natural environment, modeled by Zahorik with the function $v = -3.64 \log_2(r) + 10.76$ [19]. The two curves follow the same trajectory, suggesting that the direct-to-reverberant energy ratio in the virtual environment follows a natural behavior. Moreover the values of the ratios are much lower in the 2D mesh than in the natural auditory space as it can be noted by comparing the solid line with the dashed line in Fig. 6, which means that the amount of reverberation is exaggerated in the virtual environment.

The other known auditory distance cues, namely the Interaural Level Difference (ILD) and the spectrum, are not provided by the present model. This is motivated by their relatively weak contribution to distance perception. First ILDs, which arise due to intensity differences between the two ears, are null on the median plane and therefore will not provide any distance information for sound sources directly in front or behind the listener [14] as it is the case in our virtual environment. About the spectrum changes due to the air attenuation of high frequencies, they occur for very large distances (superior to 15 m) which lie beyond the available length of the mesh. As a result, the only pieces of information about the distance of a sound source in the mesh are the intensity and mostly the direct-to-reverberant energy ratio.

4. USER STUDY

An experiment was designed to evaluate the interface described in Section 2 in a task of audio browsing.

4.1 Method

4.1.1 Participants

Twelve Italian students (one woman and eleven men) from the University of Verona, Italy, aged 22-26 years old, participated in the experiment. All reported to have normal hearing. One is a DJ in his free time and two are musicians.

4.1.2 Stimuli and Apparatus

The playlist consisted of twelve loop samples intended for

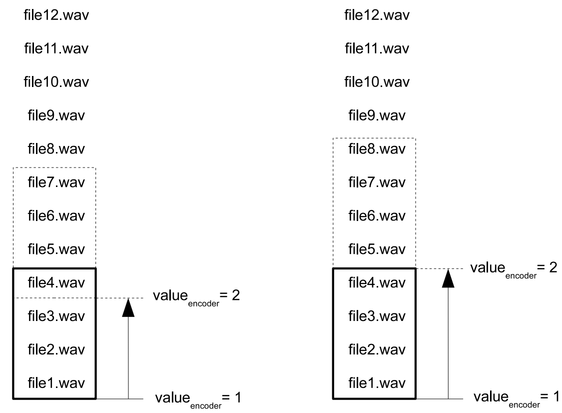


Figure 7: *Four sound files are simultaneously processed in the DWM and the increment of the knob's value enables to move the mesh 3 sound files forward (left figure) or 4 sound files forward (right figure).*

DJs found on Internet[2]. Most of them had a bpm of 130, and the others were also set at 130 bpm by time stretching. Presenting all the music pieces at the same bpm prevents from creating a cacophony in the virtual environment which makes the identification task more difficult and less pleasant. This observation was made during a preliminary user study with various pieces of pop music. In addition, we believe that the present interface may be particularly appropriate for DJs who handle loop samples and must in real time look for new ones with the same bpm. Finally, the twelve wave files were RMS equalized to avoid bias due to intensity differences between the samples.

The ribbon and the knob described in section 2 were used for the experiment and connected to the USB-ports of an Apple MacBook Pro computer. Data from the ribbon were rescaled to fit the length of the DWM. Four sound files were equally spaced on the y-axis of the 120 cm virtual membrane, placed at 6.6, 42.9, 79.2 and 115.5 cm respectively. The objective of the knob was to be able to switch the four sound files processed in the mesh and simultaneously available to the user. In a preliminary evaluation test, incrementing the value of the encoder enabled to move one sound file forward or backward in the playlist. It was however found to be not very efficient. As a consequence, it was chosen to jump over more soundfiles when incrementing the encoder's value.

Processing of the incoming MIDI data as well as spatialization in the DWM were carried out in Max/MSP. The output mono audio signal was presented over a pair of Beyerdynamic DT 770 headphones, and the sound level control of the computer was kept constant for all users. The experimental setup may be seen in Fig. 1.

4.1.3 Conditions

Advancing of 3 and 4 sound files by incrementing the knob's value were retained as the two conditions studied in the experiment and are respectively named *condition 3* and *condition 4*. Figure 7 shows how the knob's increment operates on the selection of the four concurrently playing audio files spatialized in the DWM. The reason for choosing these two conditions is to study the two different mental represen-

[2] `http://www.audiobase.com`

tations that may be induced by the two implementations. In condition 3, the overlap between consecutive windows is intended to model advancing in a linear space. This condition is therefore meant to make the user explore the whole set of sound files as a monodimensional space, which may be accessed through a window of 4 sound sources. As for condition 4, a different mental representation of the space should arise: this condition allows to find each file in only one block of four sounds, and the blocks are independent from one another. As a consequence, the mental representation is no meant to be linear since one could explore the first block, then the third one and finally the second one. In this case the different blocks are rather represented as parallel spaces containing different sound files.

4.1.4 Procedure

The two conditions were assigned to different groups of subjects. For each condition, the twelve sound files of the playlist were randomly ordered for both training and evaluation phases. After being instructed about the functioning of the two haptic interfaces, users put on the pair of headphones and were allowed to browse freely among the twelve sound files in order to get acquainted with the tool. They could spend as much time as they needed for this training phase.

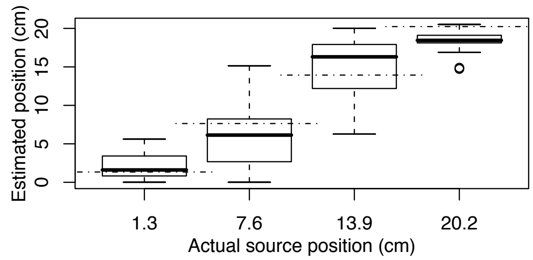
In the second phase, namely the evaluation phase, the experimenter played once a target sound file selected randomly in the playlist for 10 seconds. In the beginning of each search the knob was reset. Then the experimenter gave the go-ahead to the user for searching the sound file in the playlist. When the sound file was found, the user had to press the button above the encoder on the ReMOTE ZeRO SL. Just as the button was pressed, the time required to find the target file, the position of the user's finger on the ribbon and the value of the encoder were recorded. In addition to these data, a log file holding the values of the encoder and the position on the ribbon every 100 ms was saved for each subject. The whole experiment consisted in three consecutive sessions. For each of them, the user had to find each of the 12 sound files selected randomly among the playlist. Therefore each participant had to find a total of 36 target files.

After the experiment, a debriefing phase allowed to get a subjective evaluation of the prototype by asking users questions such as "Was the task difficult?", "What do you think about the usability of the ribbon and the slider?", "What would you suggest to improve the interface?", etc... These questions were only suggested to the users as they could freely write down their comments.

4.2 Results

4.2.1 User testing

The position of the finger on the ribbon and the level of the knob when the user reached the estimated position of the target sound file were used to compute the difference between the estimated position and the actual position of the target file. Results are displayed as a function of the actual sound file position on the ribbon. Figure 8 shows the distributions of the users' answers, for both conditions. The threshold for deciding whether the sound file found by the user was effectively the target sound file was logically chosen to be at half distance between the actual position of the sound source on the ribbon and the position of the adjacent



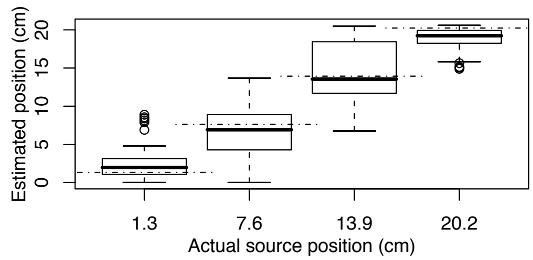


Figure 8: *Boxplots of the distributions of the estimated position versus the actual position of the target file. Outliers are marked by 'o'. Dot-dash lines represent the actual positions of the sound sources. Top: condition 3. Bottom: condition 4.*

sound source(s). In this way, the percentage of good answers could be calculated. A value of 93.02 % was achieved for condition 3, and 93.75 % for condition 4. Both conditions showing similar results, a linear regression between the estimated positions and the actual positions of the sound files was performed on all subjects. Results give a slope of 0.92 with $s.d. = 0.027$, $r^2 = 0.76$ and $p < 2 \times 10^{-16}$.

Time required to find the target sound file was also analyzed for both conditions, after taking out the wrong answers. Results are shown in Fig. 9, as a function of the position of the target file in the playlist. The Shapiro-Wilk test run on the time required to find a target file for each condition excludes that the two distributions are normal [$p < 2.2 \times 10^{-16}$ for condition 3, and $p = 1.434 \times 10^{-14}$ for condition 4]. Therefore, requirements are not fulfilled to test the null hypothesis that the two population means are equal using the Student's t-test. A non-parametric alternative is the two-sample Wilcoxon test, which reveals a statistical difference between the two conditions [$p = 0.048$]. Computation of the mean time for condition 3 and 4 leads to values of 20.47 and 15.03 respectively, meaning that in average, users under condition 4 perform the task faster than users under condition 3.

Through sessions, the average time does not change significantly under the two conditions [paired Wilcoxon tests between each pair of sessions gives $p_{min} = 0.07209$ for condition 3, and $p_{min} = 0.06445$ for condition 4]. Top and bottom plots of Fig. 10 represents the average time values for each sound file position in the playlist for the three sessions, under condition 3 and 4 respectively. Another way to investigate the evolution of the performance over the sessions is to compute average time values as a function of the position of the target sound source on the ribbon, as shown in Fig. 11. While the evolution of performance under condition 3 does not exhibit any clear behavior, it can be seen

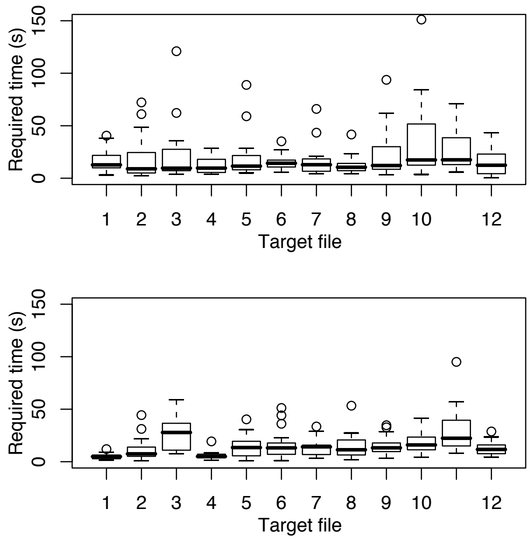


Figure 9: *Boxplots of the distributions of the time required to find the target file in the playlist. Outliers are marked by 'o'. Top: condition 3. Bottom: condition 4.*

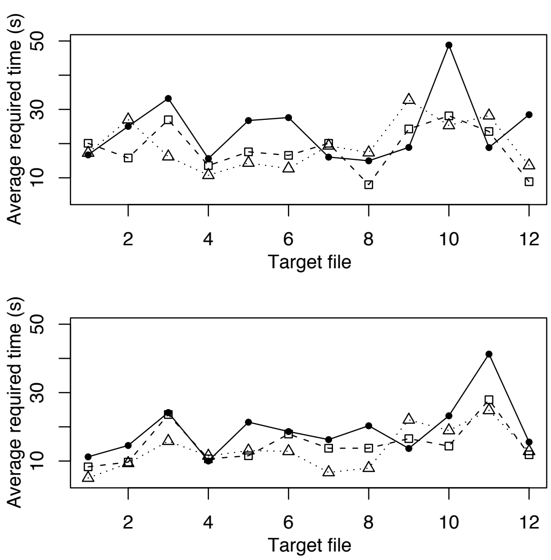


Figure 10: *Average required time to find the target file as a function of the target file's position in the playlist. Solid line: session 1, dashed line: session2, dotted line: session 3. Top: condition 3, bottom: condition 4.*

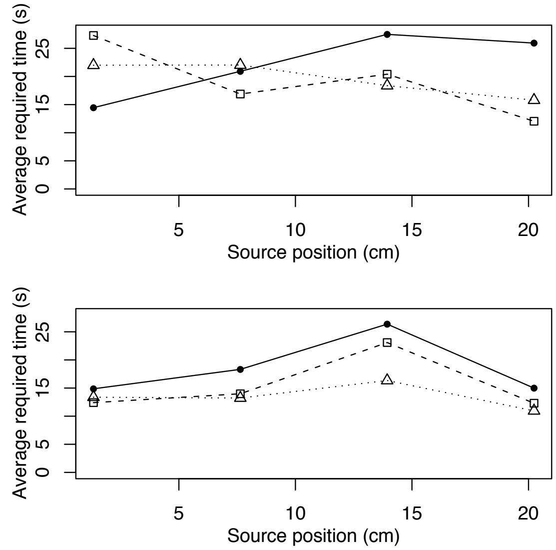


Figure 11: *Average required time to find the target file as a function of the target file's position on the ribbon. Solid line: session 1, dashed line: session2, dotted line: session 3. Top: condition 3, bottom: condition 4.*

that time decreases with training for condition 4, and in particular for the two central positions on the ribbon. At these positions, target sound files might be more difficult to find during the first trials because unlike the two extreme positions, they are surrounded by two adjoining sound sources located at equal distances from the target file.

4.2.2 *Qualitative evaluation*

All users found the interface entertaining and intuitive. They could quickly use it for searching a sound file as they reported that it was a simple and fast method for audio browsing. A few users however pointed out the confusion that could arise between some sound files, in particular among the rhythmic ones, while the task was easier for more melodic loop samples. In addition, many users suggested a more clear separation between the sound files in the virtual environment.

4.3 Discussion

From the performance evaluation, it can be seen that users manage to find the target file in the playlist using the tactile interface. Comparison between the two conditions reveals that in average the time required to reach the target sound source is smaller when the knob's increment advances of four sound files in the playlist. This result suggests that the implementation is more efficient when no sound file is repeated from one knob's position to the consecutive one. Besides, efficiency did not improve significantly through the experiment under condition 3. One could expect that some participants could remember the position of some sound files from one session to another since the playlist order remains unchanged. Although not statistically significant, results under condition 4 show a time decrease over the three sessions. The difference between the two conditions may be explained by the more simple implementation of condition 4: each sound file is played for one knob's value only, whereas condition 3 renders some files for two different values of the knob and at different positions in the virtual environment. Even if condition 3 may provide a more complex interface, the users tested under this condition did not express any particular difficulty, compared to those under condition 4. In fact, all reported that the interface was easy and enjoyable to use.

5. CONCLUSION AND FUTURE WORK

We have proposed an interface for audio browsing based on spatialization of audio data in depth. The virtual environment rendering distance cues of sound sources is modeled by a rectilinear Digital Waveguide Mesh. Associated with a rectangular ribbon playing the role of a tactile input to the interface, this coherent tool may offer new opportunities for designing auditory interfaces. Originally used as a substitute for a computer mouse, touchpads are now invading the world of portable media players and personal digital assistants. They are used as a control interface for menu navigation on all of the currently produced iPod portable music players. Another illustrative example is the iPhone that may be used as a touchpad to wirelessly control a computer. The proposed spatialization in depth of sound sources for navigation could therefore benefit from the already existing hardware and diversify the interaction modes.

An evaluation of the interface used for audio browsing was conducted. The limited length of the DWM requires another tool to navigate among numerous sound files. Then the tactile ribbon plays the role of an audio window in the virtual environment, and a knob allows to move this window along the whole audio space. A user study showed the ability of the tool to find audio files, in an entertaining and intuitive manner. The acoustic environment provided by the DWM needs however further investigations to evaluate its "naturalness" and could be compared to other techniques such as crossfading or stereo panning. Though, spatialization in depth outputs a monaural signal and therefore does not rely on the reproduction hardware configuration, unlike stereo panning as proposed by Pitt and Edwards [11]. Besides, CPU load restricted the available length of the DWM, while a more efficient computation may allow to increase the length and consequently a better sound source discrimination.

Two implementations for the knob's control on the displacement of the audio window were tested by different groups of users. A statistical analysis revealed a significant time reduction in the case of 4 sound files shift, i.e. when no audio stream is repeated from one audio window to another. While a shift of 3 sound files may induce a linear audio space representation, a shift of 4 splits the space into parallel and independnet subspaces that can be directly accessed. Under this assumption, the user under condition 4 should more easily remember which subspace the target audio file belongs to, and therefore increase its performance over time. This reasoning is supported by the time decrease with the number of sessions observed for condition 4. Further analysis may deal with the strategies used by the two groups of participants, by looking at the individual log files which provide the evolution of the audio window position and the listening position inside this window over time. Furthermore, it could be interesting to develop models for the task completion time and compare them to experimental results. For condition 4 where all the blocks are independent, the model might consist in dividing the task completion time into the time needed to find the right audio window (depending on the knob's value) and the time needed to find the target file in the audio window (depending on the position of the file on the ribbon). This model should no longer fit for condition 3 since the blocks are not independent.

The user study presented in this paper offers various research directions, both in terms of interaction design and auditory perception. Among the numerous issues, the nature and the number of concurrently playing sound sources in the audio window most probably affect the performance results. A compromise should be found between auditory overload and performance.

Finally, other uses of the interface may be explored. In particular, an auditory menu may originate from carefully sonified menu items, spatialized in depth using the DWM, and accessed through the linear tactile interface.

6. ACKNOWLEDGMENTS

We would like to thank our colleagues Stefano Delle Monache and Stefano Papetti for their valuable help, as well as Andrea Fasoli for his continuous support to the experiment. This research work has been supported by the European project FP6-NEST-29085 CLOSED - Closing the Loop of Sound Evaluation and Design.

7. REFERENCES

[1] S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proc. of the SIGCHI conference on Human factors in computing systems*, pages 473–480, 2003.

[2] M. Fernström and E. Brazil. Sonic browsing: An auditory tool for multimedia asset management. In *Proc. of the 2001 International Conference on Auditory Display*, pages 132–135, 2001.

[3] M. Fernström and C. McNamara. After direct manipulation–direct sonification. *ACM transactions on Applied Perception*, 2(4):495–499, 2005.

[4] F. Fontana. *Physics-based models for the acoustic representation of space in virtual environments*. PhD thesis, University of Verona, Italy, 2003.

[5] F. Fontana and D. Rocchesso. A physics-based approach to the presentation of acoustic depth. In *Proc. of the 2003 International Conference on Auditory Display*, pages 79–82, July 2003.

[6] A. Jensenius, R. Koehly, and M. Wanderley. Building low-cost music controllers. *LNCS*, 3902:123–129, 2006.

[7] L. Ludwig. Extending the notion of a window system to audio. *Computer*, pages 66–72, 1990.

[8] D. McGookin and S. Brewster. Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. *ACM transactions on Applied Perception*, 1(2):130–155, 2004.

[9] H. Møller, M. Sørensen, D. Hammershøi, and C. Jensen. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.*, 43:300–321, 1995.

[10] S. Nielsen. *Distance Perception in Hearing*. PhD thesis, Aalborg University, Denmark, 1991.

[11] I. Pitt and A. Edwards. Pointing in an auditory interface for blind users. In *Proc. of the 1995 IEEE*

International Conference on Systems, Man and Cybernetics, pages 280–285, 1995.

[12] A. Savidis, C. Stephanidis, A. Korte, and K. Crispien. A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction. In *Proc. of the Second annual ACM conference on Assistive technologies*, pages 117–123, 1996.

[13] C. Schmandt. Audio hallway: A virtual acoustic environment for browsing. In *Proc. of the 11^{th} annual ACM symposium on User interface software and technology*, pages 163–170, April 1998.

[14] B. Shinn-Cunningham. Learning reverberation: Considerations for spatial auditory displays. In *Proc. of the 2000 International Conference on Auditory Displays*, 2000.

[15] T. Strybel and D. Perrott. Discrimination of relative distance in the auditory modality: The success and failure of the loudness discrimination hypothesis. *J. Acoust. Soc. Am.*, 76(1):318–320, July 1984.

[16] A. Walker, S. Brewster, and S. McGookin. Diary in the sky: A spatial audio display for a mobile calendar. In *Proc. of Interaction Homme-Machine - Human Computer Interaction (IHM-HCI)*, September 2001.

[17] F. Winberg and S. Hellstrøm. Designing accessible auditory drag and drop. In *Proc. of the 2003 conference on Universal usability*, pages 152–153, 2003.

[18] P. Zahorik. Auditory display of sound source distance. In *Proc. of the 2002 International Conference on Auditory Display*, July 2002.

[19] P. Zahorik. Direct-to-reverberant energy ratio sensitivity. *J. Acoust. Soc. Am.*, 112(5):2110–2117, November 2002.