REDUNDANCY VERSUS COMPLEXITY IN AUDITORY DISPLAYS FOR OBJECT LOCALIZATION – A PILOT STUDY

Angela Constantinescu[1,2]

[1]Karlsruhe Institute of Technology,
Study Centre for the Visually Impaired Students,
Engesserstr. 4, D-76131 Karlsruhe, Germany
angela.constantinescu@kit.edu

Tanja Schultz[2]

[2]Karlsruhe Institute of Technology,
Cognitive Systems Lab,
Adenauerring 4, D-76131 Karlsruhe, Germany
tanja.schultz@kit.edu

ABSTRACT

In user interfaces, redundancy is often an indication of good design. Several studies [1,2,3], showed that when visual, haptic or other display types were combined with an auditory display, the result was an enhanced user experience and increase in performance. Research about redundancy within the auditory display alone, however, seems to be inconclusive. A pilot study was set up to test whether redundancy in auditory mappings supports object localization or rather renders the task inefficient by adding unnecessary complexity. The study used three sound parameters: pan, pitch and tempo in a combination of three sonification schemes: pan alone, pan and pitch and all three, in order to convey to the user information about the position of an object in a 2 dimensional space. Preliminary results showed that the third sonification scheme (with most redundancy) yielded the best user performance, and was also rated best by five out of seven users.

1. INTRODUCTION

In 2009, the World Health Organization (WHO) estimated that 314 million people worldwide have a visual impairment and 45 million of them are blind [4]. More than 80% of the visually impaired people are over 50 years old, so the total number is likely to increase in view of the expected population ageing. Additionally, there is a category of people who are temporarily disabled, through the nature of the task they are carrying or the environment in which they operate. An example is a firefighter in a burning building where smoke impedes all or most vision.

All these people are faced with a genuine problem when having to carry out one of the most basic tasks, both in real life as well as in computing, namely object localization. The graphical user interface can hardly be imagined without the point-and-click paradigm anymore, which implies locating a target on the screen and navigating the mouse to it. Similarly, finding objects is a crucial task in our everyday lives. Some applications for object identification and navigation based on auditory interfaces have been developed [5, 6]. However, a complete and clear conceptual framework on how to implement such an interface seems to be still missing.

One of the issues to be considered when implementing an auditory interface for object localization is whether (and to what extent) redundancy in auditory mappings either supports or overloads a user in accomplishing the task. Kramer et al. stated already in 1999 that "audio's natural integrative properties are increasingly being proven suitable for presenting high-dimensional data without creating information overload for users" [7]. In an earlier study, Kramer alone had suggested that certain redundant sound mappings and generally more complexity in sound makes the sonification "subjectively richer and easier to listen to" [8]. This suggests that complexity which may result from redundant parameter mapping should not overload the user, if the sonification is reasonably designed.

2. RELATED WORK

In various sonification scenarios, studies showed that redundancy in the auditory mappings may lead either to performance loss or performance gain. Sandor and Lane for instance [9] report that a temporal representation alone was more efficient in the mapping of absolute values to sound than both a representation with pitch and a redundant temporal and pitch representation. In a later study, Sandor [10] warns that in circumstances when precision is critical, using integral dimensions (which influence each other, such as pitch and loudness) in a redundant approach may have an undesired effect. Peres and Lane [11] on the other hand find that in the design of an auditory graph, redundant mapping of integral dimensions of sound (pitch and loudness) showed beneficial, while the redundant mapping of separable dimensions presented no gain. Yet an older study [12] seems to point to an increased performance measured in response time, when a combined pitch and pan mapping was used redundantly than when the two sound parameters were used individually.

From the work presented, no clear conclusion can be drawn about the conditions under which redundancy is useful or not: are certain redundant combinations of sound parameters beneficial while others are generally not? If so, does integrality of the parameters play a role? Moreover, are certain data types or circumstances better represented through redundancies?

This study showed that in the context of object localization in a two dimensional space, the combination of pan, pitch and tempo used to convey redundant location information was more efficient and subjectively better rated than pan alone or pan and pitch. The three sound parameters were not mapped to the same data values, but instead each was mapped to give a different location clue: pan for x axis deviation, pitch for y axis deviation and tempo for distance to the object. However, the combination of two or three mappings is considered here to be redundant because the pan mapping alone is sufficient to allow the user to find the object, and adding an extra mapping only gives the user more strategies to choose from as well as more feedback, but it does not inherently change the initial strategies. The study also showed that five out of seven participants followed the same strategy to find the object when all three sound parameters were mapped as in the case when either only one or two mappings were in place.

3. EXPERIMENTAL SETUP

The participants in this study were required to play an audio game that simulated object location in a two dimensional space. A visual component was implemented for demo as well as for logging purposes. The game consisted of one small square (the object) placed at a fixed position on a rectangular canvas (the room). The user could navigate in the room using the computer mouse, which at the beginning of the game was placed at the bottom, centered. Mouse control was chosen over key control because movement with the mouse simulates better the lack of precise orientation that a blind person would have in such a scenario. Over a set of headphones, the user could hear the sonification that described the position of the object in relation to the mouse cursor (the person). In the room, random obstacles may have been placed. These had all rectangular shapes and kept the user from advancing, forcing him or her to move around them.

There were three sonification strategies used: in the first one, only pan varied with displacement on the x axis, while pitch and tempo were constant. In the second strategy, pan and pitch varied with displacement on the x and y axes respectively, while tempo remained constant. In the third strategy, pan and pitch varied as in the second strategy, while tempo was given by the distance between cursor and object.

Each sonification strategy was repeated six times, resulting in a total of 18 trials. In the first three trials of each strategy there were no obstacles in the room, while in the next three trials three obstacles were placed at various locations between the initial cursor position and the object in order to maximize the probability that at least one obstacle was encountered. In each trial the object was placed at different locations, but always at the same distance from the initial cursor position. The goal of each trial was to find the object, that is, to place the mouse cursor over the small square.

The experiment took between 45 minutes and two hours to complete and was carried out by six out of seven participants in one session, while one participant interrupted the experiment and continued the next day. The same hardware equipment was used for all participants.

3.1. Auditory display

The sonification techniques used in this experiment were parameter mapping and earcons. Piano sounds were played intermittently in a sonar-like manner. When the user hit a room wall, a stifled hitting sound was simulated by an electric muted guitar. When the object was reached, the success sound that informed the user of the task completion was played as an F6 on the xylophone. Pan, pitch and tempo have been chosen to map x axis, y axis and distance respectively.

The choice of sound parameters reflects the most widely used and successful mappings found in applications today: pan for x axis deviation is used for instance in audio games; a lot of research is now conducted in the area of 3D sound, since sound direction is one of the most intuitive mappings in localization or navigation tasks. The choice of pitch for the y axis was partially based on the Doppler Effect but also considering applications specifically designed for blind users such as the Audio Graphing Calculator from ViewPlus Technologies [13]. The mapping of distance to tempo follows the sonar and the car parking system paradigms, with which many people nowadays should be familiar.

Here is a description of the auditory mappings used:

Pan: the difference on the x axis between object and cursor mapped to pan. The object was viewed as the sound source: when the object was left of the cursor, the sound came from the left and when the object was right of the cursor, the sound came from the right. In order to avoid ambiguity and to strengthen the perceived effect, only marginal pan values have been used.

Pitch: the highest value for pitch was played when the cursor was aligned with the object on the x-axis. As the cursor moved above or below this line, the pitch went down.

Tempo: as the cursor approached the object, the tempo of the sound increased

Percussion: in order to avoid overshooting due to delays or perception errors, a percussion sound was played on the object's y axis, that is, when the cursor was aligned under or above the object, similar to what Caperna et al. [6] have used in their navigation system for blind people.

Bass: just like percussion for the x axis, a bass sound was added to the sonification in order to support the user in the orientation on the y axis.

Finally, in the first sonification strategy, when pan alone (together with the sustaining percussion) was used, no sound was heard above the object. This is one way to replace the missing bass orientation, and also to provide a parallel to a real-case scenario by simulating a situation where the object is out of the vision field.

Here are the six stages that each participant had to complete, in this order (two stages for each strategy, with and without obstacles):

p: pan

p-o: pan with obstacles

pp: pan and pitch

pp-o: pan and pitch with obstacles

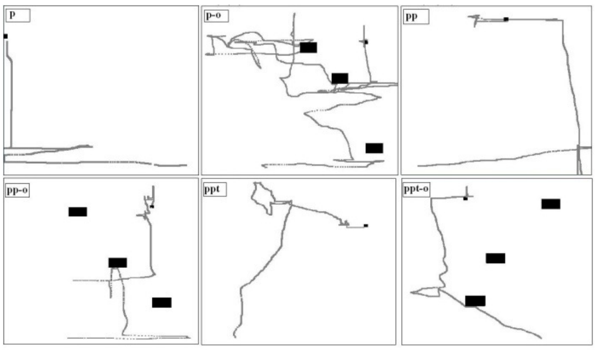
ppt: pan, pitch and tempo

ppt-o: pan, pitch and tempo with obstacles.

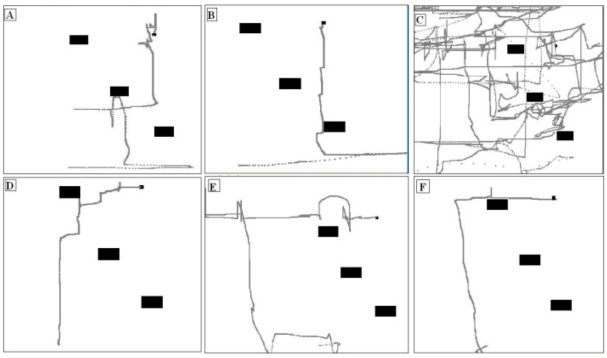
3.2. Participants

Seven users, three men and four women, with ages between 34 and 48 participated in the experiment. Their computer skills ranged from medium user to expert, the orientation skills from very poor (cannot read maps and gets lost often) to very good (can read maps very well and almost never gets lost) and five out of seven played an instrument. One user was blind from birth and another one was late blind.

This group of participants is quite heterogeneous, and moreover, it is important to note that unlike in many experiments reported in the literature, the participants are not undergraduate students with a possibly strong technical background. I consider this to be an advantage, since the participants that took part in this study represent better the target users of a possible navigational system, who have the most various skills and knowledge.



(a) Participant A, one trial from each stage



(b) Stage pp-o, participants A-F

Figure 1: Screenshots of the mouse trail

The participants had no prior experience with the system being tested. A small demo was held just before testing as well as before stages p, pp and ppt in order to explain each sonification strategy. For blind users the demo was accompanied by spoken descriptions of what could be seen on the screen.

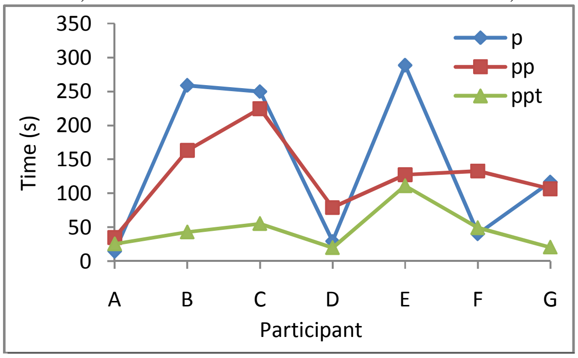
3.3. Data collection

The success rate of each trial was recorded; no time limit was given for the experiment; however, the participants had the option to give up a trial whenever they considered necessary. The times to complete each task have also been logged (see Table 1), as well as screenshots of the cursor trail at the end of each trial (see Figure 1). The subjects were observed throughout the whole experiment duration by a supervisor, who documented user comments, behavior and solution strategy.

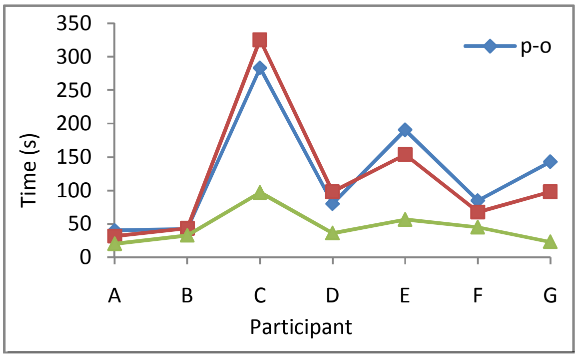
4. RESULTS

4.1. Differences between strategies

The average completion times of all participants for each stage (see Figure 2 and Table 1) show that the last two stages, where pan, pitch and tempo were used together in a more complex sonification account for the smallest completion times. This is also consistent with the users' preference for the last sonification strategy: five out of seven participants liked this strategy best, and also (other) five considered it is the most intuitive, while one user could not decide. For users A, D and F,



(a) No obstacles

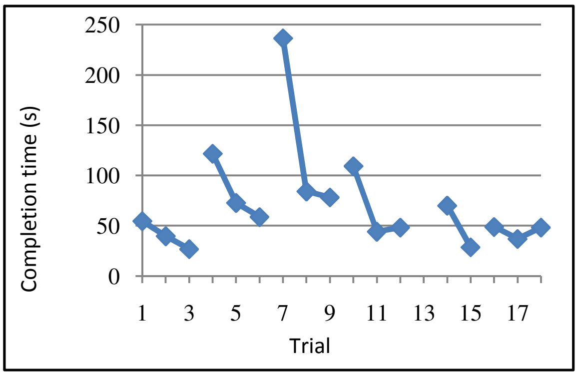


(b) With obstacles

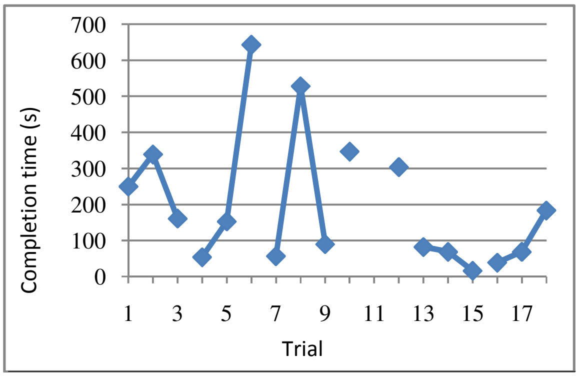
Figure 2: Average completion times over trial for each stage, for all participants

two of which were the blind users, the differences in the average completion times between the different stages were minimal (see Figure 2). All of these three users have still reported though that their favorite strategy was the last one.

The average completion times do not seem to differ much for the first two stages. This implies that the redundancy in the second sonification (mapping the y axis to pitch) did not bring any improvement, while the redundancy in the third sonification strategy (mapping distance to tempo) did. This might be explained by the fact that the second sonification adds one more localization strategy (namely go up until reaching the highest pitch then follow left or right) to the existing one (go left or right until reaching the percussion, then go up), which is, however, discrepant as the two do not influence each other. At best, the pitch might give the user additional feedback when pursuing the location strategy that follows the percussion. The tempo in the third strategy on the other hand gives a quick continuous feedback which is harder to ignore than the change in pitch or pan, no matter which of the two strategies the user was pursuing. Additionally, some participants went slightly or



(a) Participant for whom the learning effect is most obvious



(b) Participant for whom no learning effect is visible

Figure 3: Completion times for two participants: the learning effect is visible only for the first one

directly on a diagonal path in the last two stages (last sonification strategy), which might also account for smaller completion times.

There are no conclusive differences in completion times between stages without obstacles and stages with obstacles, although three users said that they found stages with obstacles particularly challenging.

4.2. Differences between trials: the learning effect

A learning effect could only be clearly identified for two out of seven participants. Figure 3 shows two examples, a) where the times within one stage go down from one trial to the other, the learning effect being thus the most obvious and b) where a learning effect can only be assumed in one out of six stages, namely the fifth. Another point that indicates that little, if any learning at all has taken place is the fact that one participant took once more a trial from stage p-o, after she had completed the entire experiment. The time measured for this last test was greater than the average for stage p-o and was comparable to the first and largest time in that stage.

Since no learning effect could be markedly recognized, one may assume that the lower times recorded in the last strategy (ppt and ppt-o) were not an effect of practice and that indeed this strategy was the most efficient.

4.3. Differences between participants

Walker and Mauney [14] ran a study to assess the relationship between individual differences and the interpretation of auditory displays and more specifically graphs. They found out that working memory and gender are very likely to be good predictors of sonification interpretation, while other individual characteristics such as age, handedness or musical ability do not seem to be good predictors or at least that no conclusive statement could be drawn and more research should be conducted. The current study reports a possible correlation between age and handedness and mapping preferences in auditory display.

At least three of the participants in this study considered it more intuitive to move the mouse to the left when the sound came from the right, as though running away from the sound. This might make sense in situations when the sound is used to alert the user of a wrong move, for instance. Two of these participants also mentioned that they are left-handed. Of course, no conclusion can be drawn from this information alone, however, it might be sensible to further investigate this account.

A more important discovery however was that two out of seven participants had trouble in distinguishing the panning effect. It is maybe by chance or maybe not that these persons were the oldest in the group, with ages approaching and respectively exceeding 45. It is crucial to further investigate whether this high number was just a coincidence and also whether this implies that 3D sound perception is indisputably affected as well. Recent research has been focusing on 3D sound simulated in stereo headphones using HRTFs. If it is proven that many people, whether due to age or other reasons, do not have the ability to interpret 3D sound, then alternative auditory displays should be researched instead or in parallel with 3D audio.

Finally, the participant who was blind from birth made the suggestion that for visually impaired people, it would make sense that the mouse be also sonified. This is due to the different perception that blind users have over space: while sighted people experience the world from the point of view determined by their current location and thus view objects around them in relation to themselves, blind people must place both objects and themselves in the world (room, street, etc). Another pilot study should be conducted in order to test such a design.

No other differences in sonification preferences have been noted between the visually impaired and the sighted users, although accounts of such disparities exist in the literature [15, 16].

5. CONCLUSION

This study analyzed three sonification strategies in the context of a 2D navigational task. The first sonification maps x axis to pan, the second adds the mapping of y axis to pitch and the third adds a third mapping of distance to tempo. It is stated that the second and third strategies include redundant information, since the first mapping alone allows the user to construct a localization strategy and to navigate to the object. It is shown that while the second strategy does not seem to bring any performance gain, the third (mapping distance to tempo) results in better completion times and is better rated by the users. It is believed that the smaller average completion times in the third strategy do not result from practice.

Part.		***1***	***2***	***3***	***4***	***5***	***6***	***7***	***Avg***
p	***1***	10	598	250	30	250	55	175	195
	2	16	110	339	28	66	40	58	94
	3	19	69	161	30	550	27		143
Avg		15	259	250	30	289	40	117	143
p-o	***1***	5	74	54	173	171	122	143	106
	2	83	31	153	29	188	73		93
	3	32	22	643	37	213	59		168
Avg		40	42	283	80	191	85	143	123
pp	***1***	51	191	57	156	160	236	137	141
	2	33	54	528	48	148	84	95	141
	3	22	245	90	34	75	78	89	90
Avg		35	163	225	79	128	133	107	124
pp-o	***1***	15	47	347	126	288	109	89	146
	2	47	45		41	134	44	54	61
	3	32	38	304	126	38	48	151	105
Avg		31	43	325	98	153	67	98	117
ppt	***1***	21	34	82	25	131		19	52
	2	19	17	69	17	162	70	14	52
	3	37	79	16	18	41	29	30	36
Avg		26	43	56	20	111	49	21	47
ppt	***1***	24	11	39	30	29	49	19	29
-o	***2***	20	47	69	30	80	37	24	44
	3	16	40	184	48	60	48	26	60
Avg		20	33	97	36	56	45	23	44

Table 1: Completion times and averages in seconds for all participants, over all trials

The study also points to possible relationships between individual differences and auditory mapping preferences, such as pan mapping polarity which could be related to left-handedness, and reports that pan discrimination was an issue in two out of seven participants.

6. REFERENCES

[1] Z. Zhou, A. D. Cheok, X. Yang, Y. Qiu, “An experimental study on the role of 3D sound in augmented reality environment,” *Interacting with Computers,* 16, pp. 1043–1068, 2004

[2] W.W. Gaver, R.B. Smith, and T. O’Sheak, “Effective sounds in complex systems: the ARKola simulation,” in *Proc. of CHI ’91*, 1991, pp. 85–90.

[3] M.J. Bishop, and W. M. Cates, “Theoretical Foundations for Sound’s Use in Multimedia Instruction to Enhance Learning,” *ETR&D*, vol. 49, no. 3, pp. 5–22, 2001.

[4] http://www.who.int/mediacentre/factsheets/fs282/en

[5] C. Magnusson, and K. Rassmus-Gröhn, “Audio haptic tools for navigation in non visual environments,” in *Proc. of the 2nd Int. Conf. on Enactive Interfaces (ENACTIVE05)*, Genoa, Italy, 2005.

[6] S. Caperna et al., *A Navigation and Object Location Device for the Blind*, PhD thesis, University of Maryland, College Park, 2009.

[7] http://icad.org/node/400

[8] G. Kramer, "Mapping a single data stream to multiple auditory variables: A subjective approach to creating a compelling design," in *Proc. of the Int. Conf. on Auditory Displays*, Palo Alto, California, USA, 1996.

[9] A.Sándor and D. M. Lane, "Sonification of absolute values with single and multiple dimensions," in Proc. *of the Int. Conf. on Auditory Displays*, Boston, MA, 2003.

[10] A. Sándor, *Perceptual interaction of duration with pitch and rate of change in pitch: Implications for sonification.* 2004, Rice University: Houston, Texas

[11] S. C. Peres, and D. M. Lane, “Auditory graphs: the effects of redundant dimensions and divided attention,” in *Proc. of the 11th Meeting of the Int. Conf. on Auditory Display (ICAD05)*, Limerick, Ireland, 2005.

[12] J. R. Simon, and J. L. Craft, “Communicating directional information with an auditory display,” *http://www.apa.org/journals/apl.html*, vol. 55, no. 3, pp. 241-243, 1971.

[13] C. A. Supalo, T. E. Mallouk, C. Amorosi, L. Rankel, H. D. Wohlers, A. Roth, and A. Greenberg, “Talking Tools to Assist Students Who are Blind in Laboratory Courses,” *J. of Science Education for Students with Disabilities,* vol. 12, no. 1-Winter, pp. 27-32, 2007.

[14] B. N. Walker, and L. M. Mauney, “Individual differences, cognitive abilities, and the interpretation of auditory displays”, in *Proc. of the 10th Int. Conf. on Auditory Display (ICAD04)*, Sydney, Australia, 2004.

[15] S. W. Mereu, and R. Kazman, “Audio Enhanced 3D Interfaces for Visually Impaired Users,” in *Proc. of CHI 96*, Vancouver, Canada, 1996.

[16] B.N. Walker, and L. M. Mauney, “Universal Design of Auditory Graphs: A Comparison of Sonification Mappings for Visually Impaired and Sighted Listeners,” *ACM Trans. Access. Comput.* 2, 3, Article 12. 2010.