Directing Untrained Users' Attention Using Simple Sound Patterns

Lars Bo Larsen[1(✉)], Søren Staal Jensen[2], and Mai Baunstrup[3]

[1] Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark
lbl@es.aau.dk
[2] Timeplan, Vandmanden 10C, 9200 Aalborg, Denmark
sjensen1984@gmail.com
[3] Visma, Karenlyst Alle 56, 0277 Oslo, Norway
baunstrup@gmail.com

Abstract. In this work, we show how simple sound patterns can be used to direct a person's attention towards specific locations, when the sound is emitted from a single, fixed position. In interaction design, this is an effect most often achieved using visual ques, such as animations or flashing lights. Using sound instead is useful in situations where visual ques are undesirable or unfeasible. We designed two experiments with a total of n = 45 respondents. 10 different sound patterns were composed with inspiration from previous studies and used as stimuli. The results showed significant differences (p = 0.05) for directing the gaze upwards and downwards and trends for left-right. The composed sounds require no prior training, or knowledge of cultural references (contrary to e.g. earcons and auditory icons) by the respondents and can thus be regarded as universal in nature.

Keywords: Human factors · Interaction design · Auditory displays · Sound design · Directing attention

1 Introduction

Human interaction with technology today is highly based on displays and the use of the visual sense. However, there are different social and work situations in which a flashing or dynamically changing display would be inappropriate. This could be either in accordance to social acceptability or in situations where the visual senses are occupied elsewhere, such as automotive contexts. Furthermore, not all devices have a visual display. Consequently, there has also been much research into the field of using sound in interaction with technology, such as sound scape design, virtual 3D sound, use of earcons and auditory icons. In the work presented here we are interested in replacing the visual modality with the auditory in certain situations.

A common problem in interaction design is how to capture and direct the users' attention towards a particular location, e.g. on a screen or device front plate. A simple and well-used solution is to use dynamic visual elements, such as a blinking LED on a device front plate or an animated icon, as known from common OS GUIs.

A.G. Ho (ed.), *Advances in Communication of Design*, Advances in Intelligent Systems and Computing 609, DOI 10.1007/978-3-319-60477-0_2

In the present work, we focus on the ability of sound to solve this problem. The solution corresponding to the visual scenario described above would be to use directed sound, either emitted from a particular physical location or a virtually generated 3D sound, to draw the users' attention towards the sound source. This is further discussed in Sect. 2.

However, this solution imposes some undesirable constraints and requirements on the design-space, both seen from the designer's and the manufacturer's viewpoint. Regardless whether real or virtual localized sound is used, the scenario calls for a number of speakers to be located at strategic positions on or around the device. The user will also be required to be located in a particular position or area relative to the device emitting the sound. This is necessary to correctly perceive the sound source location. This will make it more difficult to achieve a good design, restrict the number of users to one at a given time and will make production of a device more complex and expensive compared to a single-speaker solution.

Discarding both the real and virtual localized sound solutions, we will instead turn towards investigating the design of particular sound patterns. In other words, we will use the sound's intrinsic characteristics instead of the localization of the sound in an attempt to direct the user's spatial attention. This is similar to what is sought by earcons and auditory icons, except here we are looking strictly for the *perceived spatial characteristic*. We expect this will introduce certain constraints regarding e.g. the nature and duration of the designed sounds, but will nullify the need for specific spatial placement and number of speakers, as well as the users' position.

As a brief outline of the paper, the next section describes related work in the field of designed sounds, which can guide the user in certain directions. Following this, an experiment with specific sound patterns will be described. Finally, the results and discussion are presented together with the conclusions drawn from the study.

2 Related Work

One of the first important studies in this field is from 1930. In an experiment, Pratt placed loudspeakers behind a screen at five different height intervals with a fourteen-point scale in front of it [1]. The loudspeakers emitted tones at 256, 512, 1024, 2048 and 4096 Hz. Pratt then played the tones back in randomized order and location. The subjects were told to indicate where they thought the sound came from. The result of the experiment was that high frequency tones were generally perceived higher up in space than the tones of lower frequency, regardless of the physical location the tones were emitted from [1, 2].

In a later experiment Roffler and Butler (1968) placed four loudspeakers on top of each other behind a screen divided into thirteen numbered sections. During the experiment test-subjects sat 1.5 meters from the screen, and they were told to indicate where they thought a sound originated from, via a number on the scale. The sounds used were pure sine tones and varied from 250 to 7200 Hz, with a duration of 10 ms. The results were similar to those found by Pratt [2, 3]. Other researchers have reported similar findings [4, 5].

More recently (2000), Walker and Ehrenstein [6] carried out experiments, in which they explored whether sound with an increase or decrease in pitch, could be perceived as having a direction. Twelve auditory stimuli were produced in the ranges 316 Hz to 2512 Hz. The auditory stimuli all had an onset pitch and then either increased or decreased in pitch to indicate either an "up" or "down" sound pattern. All sounds had durations of 150 ms. The accuracy of the responses was greatest (96%) when the pitch change was congruent, in the sense that the pitch either started high and increased or started low and decreased. This was compared to the 87% accuracy of incongruent pitch change [6].

One of the few examples of successful designed sounds, that guided people left or right was done in a study by Walker in 1987 [7]. 828 subjects were tested to see how they matched acoustic features (frequency, waveform, amplitude, and duration) to visual metaphors on a response sheet. The subjects were asked to listen to sounds designed in accordance to the mentioned acoustic features. The subjects were then told to match these sounds to either: shape, size, pattern, vertical or horizontal (via choices on the response sheet). Sounds which were based on duration were consequently matched with horizontal (i.e. left-right), and frequency with vertical (up-down). The duration sounds were designed by either using a pure or complex tone, with the frequency kept constant [7].

Thus, research into the perceived spatial location of sound patterns has been ongoing for many years. Most of these studies were designed with the presupposition that a certain sound should be associated with a location or even a metaphor presented visually. Thereby, the subjects were primed with a certain level of prior expectation about matching sounds to locations, or directing them to push certain areas to map a location for the sounds. Other studies contain acoustic experiments explaining guidelines, with distances and angles, from where a person can locate the position of the actual sound source [2].

2.1 Use of Sound Patterns in Interaction Design

William Gaver (1986, 1989) discussed how sounds are interpreted by humans and how this can be exploited in interaction design [8, 9]. In particular, he introduced the concept *auditory icons*. Gaver focused on the interpretation of the sound pattern, in contrast to the studies mentioned above, which focused primarily on the psychoacoustic perception of the sound pattern and not the semantics a sound may convey to humans.

Auditory Icons. Gaver defined auditory icons as "*caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes*" [8, p. 169]. The principle is similar to visual icons and relies on the users' ability to map the sound from a physical world to the semantics of the auditory icon. In other words, the use of auditory icons relies on a certain level of cognitive effort by the user. Either as a conscious mapping of unlearned auditory icons or a recall of already learned ones [8, 9].

Earcons. Blattner (1989) and later Brewster (1994) investigated earcons [10, 11]. Earcons are small musical messages. The building-blocks of earcons are motives, which are short melodies used as individual recognizable entities. Using rhythm and different combinations of the same motive it is possible to create a whole family of earcons, which has its own understandable hierarchy. These can be efficient and highly recognisable (such as the MS Windows sign-on jingle or error message). However, there are also certain disadvantages. First of all, earcons are based on abstract musical messages, with no mapping to real-world sounds. This means that the user has to learn and retain the meaning of them. They cannot use their knowledge of the world in order to figure out what a given earcon should represent, unlike auditory icons.

A combination of auditory icons and earcons has also been described [9]. This is achieved by modifying auditory icons according to the principles pertaining earcons.

It is important to note that both auditory icons and earcons rely on a semantic interpretation of the sound content. Initially, when learning the semantics (especially true for earcons) and subsequently, when they are encountered in auditory interfaces. Thus, there is a certain cognitive effort associated with both earcons and auditory icons. Auditory icons might need less learning and recall, but are more context dependent and studies have indicated that earcons are perceived as more pleasant [12].

2.2 Goals of the Present Study

Having briefly outlined how humans spatially perceive certain sound patterns and how the semantics of sounds can be characterized and used in interaction design, we will now describe the goals and motivations behind the present study.

The use of multiple sound sources (loudspeakers) to create real or simulated sound source locations are discarded due to the drawbacks imposed by the positional requirements described above. We instead turn to exploiting the semantic content of the sounds. However, while it has been shown that a spatial message to some degree can be encoded in auditory icons they still impose a certain cognitive load on the user, when decoding the meaning. Furthermore, they require the listener to have a specific prior experience (or a specific enrollment session) and maybe even a certain cultural background to enable decoding.

The aim of this study is therefore to explore whether a sound pattern can guide a person's attention in a certain direction, without any prior contextual knowledge or training and with a single sound source. By this, we aim to exploit low-level cognitive processing and eliminate any learning or recognition requirements. This requirement would exclude single-pitch tones, such as the studies reported in [1, 2] as these are based on comparisons between the tones and thus require prior exposure to the sounds. However, we chose to include single pitch tones in the study for comparison to the other stimuli in the experiment described in Sect. 3.4 below.

We carry out the study through an experiment, where the respondents perceived location of the emitted sound patterns is observed and recorded. Their subjective experiences are obtained by a post-test questionnaire and interview. The experiment is described in Sect. 3 below and the results in Sect. 4. In the following we will denote the sound patterns as D-sounds to indicate the intended directional capabilities.

3 Experimental Approach

We design an experiment to explore, if certain sound patterns (D-sounds) can guide an untrained persons' attention to a certain location or area on a surface, independent of the actual sound source localization. Furthermore, we also want to find out if the context of the D-sounds can override the acoustical directions from where the sounds are actually coming from. The experiment is divided into two parts, A and B as described below.

3.1 Description of the Experimental Setup

The experiment utilises a custom-built device consisting of sixteen loudspeakers arranged in a 4 × 4 grid. The loudspeakers are hidden from the respondents behind a grey cloth screen. The device is shown in Fig. 1. The respondent is placed in an adjustable chair one meter in front of the device with her head centred at the same height as the centre of the speaker device.

Ten different D-sound patterns (see Sect. 3.4 below) were played at a normal listening level in random order, and the respondent was asked to point at the speaker device with a laser pen at the location, where she perceived the sounds were emitted from. 64 sounds were played in a row. After each series of 64 sounds the respondents

Fig. 1. Left: loudspeaker device used in the experiment. Top right: close-up of the 16 individual speakers without the cover. Bottom right: facilitator with computer and respondent pointing with the laser at the speaker device (laser dot at arrow)

were given a small break to prevent overstimulation and fatigue. A questionnaire was handed out right after the experiment and a short explorative interview was conducted for each respondent.

3.2 Respondents

45 adults participated in the experiment. 28 were male and 17 female, aged from 18 to 61. Everyone reported to have normal hearing and not experiencing listening problems. We did not submit the respondents to an audiometric screening test, as the task did not include a broad range of frequencies and the loudness was deemed sufficient for persons without hearing impairments. The sound level could be adjusted, if a respondent requested it, as this was not an experiment variable, but nobody did.

3.3 Materials

The equipment used in the experiment consisted of: A custom built array of 16 2” loudspeakers arranged in a four by four grid (see Fig. 1), roughly measuring 45 × 45 cm, a custom built multi-channel Bang and Olufsen Ice Power amplifier, a TC Electronics pre-amplifier, controlled via a MATLAB program running on a MacBook Pro. This was also used for data recording and analysis.

3.4 Stimuli

The independent variables are the D-sound pattern stimulus and the speakers emitting the sound. A musician composed ten sound patterns. They consist of simple piano tones with a duration from 0.9–1.7 s and pitch varying from 100–900 Hz. These are:

- Two patterns with increasing or decreasing tempo, inspired by [7].
- Two hybrid patterns, with both tempo and pitch change.
- Two pitch change patterns, inspired on the results reported in [6].
- Four patterns with high, medium high, medium low and low pitch, inspired by [1].

Table 1 below shows the characteristics of the D-Tones

The hybrid sounds were included to investigate whether it would be possible to make a “diagonal” sound pattern, based on the up and down sounds of pitch change and the left and right sounds of duration. Five different loudspeaker configurations were used in order to facilitate the control and the experimental conditions, see Fig. 2 below. Four of the five loudspeaker configurations were used as a control condition: Here the eight loudspeakers in either the upper, lower, left or right half of the 4 × 4 loudspeaker grid were used. For the experimental condition, all sixteen loudspeakers were active at the same time. The 10 D-sounds were played 2 times across each of the 4 control loudspeaker configurations. The D-sounds were played 8 times across experimental configuration. The control condition was included in order to reveal whether the D-sound patterns could override the acoustics in the users’ response.

Table 1. The 10 D-sounds used in the experiment.

Name	Duration (Sec)	Tones	Frequency (Hz)	Tempo (Bpm)
Tempo up	1.7	8	444	250–444
Tempo down	1.7	8	444	440–180
Diagonal up	1.7	10	109–882	191–666
Diagonal down	1.7	10	882–129	740–192
Appagio up	1.5	10	139–887	400
Appagio down	1.5	10	887–142	400
High 2	0.9	1	882	–
High 1	0.9	1	444	–
Low 2	0.9	1	221	–
Low 1	0.9	1	107	–

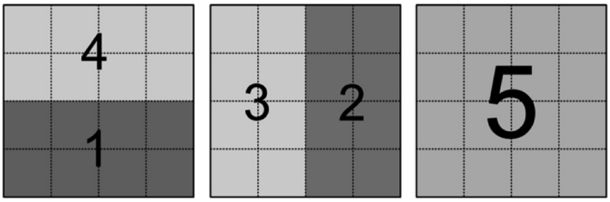


Fig. 2. Areas 1–4 indicate the bottom, right, left and top speaker subsets. Area 5 indicates all speakers are active

3.5 Procedure

The respondents were introduced verbally to the experiment and asked to sign a consent form. The experiment was split into two parts, A and B. 30 respondents participated in part A, where the D-sounds were either rendered by all the speakers, or a subset of them. See Fig. 2 above. The purpose of experiment A to verify whether the respondents were capable of identifying the location from which the sounds were emitted and how this influences the D-sound effects. These were emitted either from a specific area (1–4) on the speaker device using a subset of the loudspeakers, or from all of the loudspeakers simultaneously (5). The stimuli were presented in random order.

15 new respondents participated in part B, where all loudspeakers (5) were active throughout and the D-sounds were emitted uniformly across the surface of the device. Thus, the only spatial clues present are due to the sounds themselves. Every time a sound is played, the respondent would use the laser pointer to indicate where s/he perceived the sounds' location to be. This is observed by the facilitator and entered on the experiment PC by clicking in a 16 × 16 grid shown on the screen. As soon as the entry is recorded, the next sound is played.

4 Results

Overall, part A of the experiment shows that, when a subset of speakers is used (area 1–4), the respondents were indeed able to detect this and identify the locations reliably. However, the individual sounds also played a role. We use heat maps corresponding to the 4 × 4 response grid to depict the results.

Figure 3 shows an example of the responses to the "Appagio Down" or decreasing pitch D-sound. Comparing the responses from areas "1" and "4" shows the combined effects of the sound as well as the position. When the D-sound is only emitted in the upper area (4), respondents still perceive it as mainly coming from the bottom area, but with a tendency towards the middle of the device. The heatmap for area (5) confirms this. Area (2) right, and (3), left, shows the same tendencies, although respondents clearly are able to detect from which side the sound is emitted. This is confirmed in the exit interview; without being asked directly, many of the participants expressed that they quickly learned to distinguish between the spatial locations of the sound emissions to make their response.

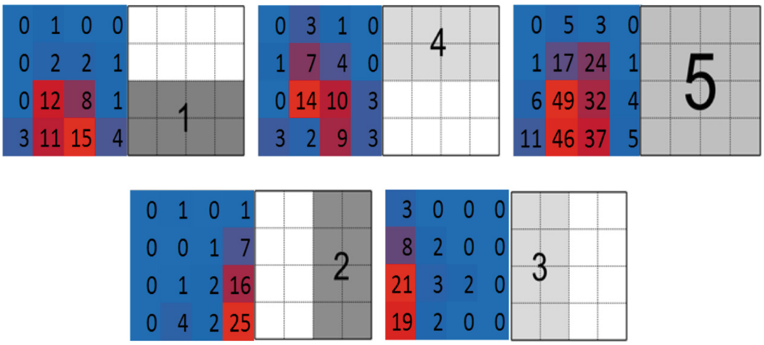
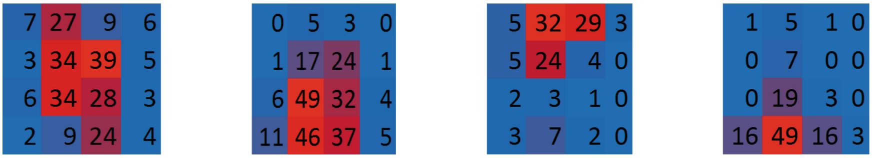


Fig. 3. Heatmaps for "Appagio Down" (Decreasing Pitch) for areas 1–5 for Experiment A.

We saw a similar tendency for the increasing pitch D-sound and the high, mid-high, mid-low and low static pitch D-sounds. However, the D-sounds with tempo changes and the combination D-sounds did not yield any clear results. Thus, even though experiment A confirmed the ability to direct the perceptual experience of the sound source location, it also documented that a habituation, or learning effect had occurred and no clear conclusions were made.

In order to eliminate the impact of the physical location, as well as the observed learning effect, all speakers were active in Experiment B. Thus, the respondents did not get any spatial information to base their responses on. In this case, CHI-square tests did indeed yield significant ($p = 0.05$) results for the decreasing and increasing pitch changes (Appagio Up and Down), and the static high, medium-high and low pitch D-sounds. The D-sounds with increasing or decreasing beats, intended for left-right directional effect, as well as the combined sounds indicated some trends, but did not yield significant results.

85% of the responses for the "Increasing Pitch" D-sound were situated in the upper half of grid and 88% of the "Decreasing Pitch" responses were situated in the lower half. Figure 4 below compares the heat maps along the corresponding Heatmaps from experiment A.



Experiment A. Experiment B.

Fig. 4. Heatmaps for increasing (left) and decreasing (right) pitch for experiment A and B for area (5).

When asked, about the right-left effects in the exit interview, half the respondents (53%) confirmed they had experienced the effect, despite non-significant results in the observed data. The corresponding results for the up-down effects were 100% for "up" and 93% for "down".

The impact of the learning effects in Experiment A is evident when comparing the Heatmaps from the two experiments.

5 Discussion

The results from Experiment A indicate an effect of the D-sounds, even when the respondent is aware that the sounds are emitted from different locations on the speaker device. The effect was strongest for up-down direction, whereas right-left direction was less pronounced. However, no significant results were obtained in this experiment. This might be due to a learning effect, where respondents became aware of the different speaker configurations (1–5) and this influenced the results.

Experiment B was carried out with all 16 speakers active at all times. Thus, any observed effects can only be due to the respondents' perception of D-sounds. In this case, significant results were obtained for the up-down direction, but not the right-left direction. We ascribe the learning effects to the different results obtained in Experiment A and B.

5.1 Impact of Chosen Sounds

Clearly, the specific design of the D-sounds must have an impact. In the present case, piano tones were used as they were considered to be neutral in the sense that they are not used as auditory icons or earcons in the most common computer systems. Furthermore, they are considered pleasant to listen to. The individual notes were clearly distinguishable. The duration of the D-sounds was up till 1.7 s. This is fairly high compared to the studies mentioned in Sect. 2. If the sounds were to be used in an interactive system, the duration will most likely need to be reduced to less than half a second.

No clear results were achieved concerning the increasing and decreasing tempo, intended to induce a right or left direction. This effect is obviously harder to achieve with a simple pattern and may even be culturally dependent on e.g. the reading direction of one's culture. We asked all respondents, if they associated the sounds with a direction and about 50% replied that they did. One respondent replied that on a piano the keys are laid out with low tones to the left and high to the right. Thus, an increasing pitch indicated a left-to-right direction to her. However, this was the only comment we received regarding this issue.

5.2 Impact of Observation Method

The location of the laser dot used to indicate the respondents perceived location of the D-sound was recorded by the facilitator observing and noting which of the 16 sub-quadrants on the speaker device the dot appeared. There is some uncertainty associated with this. However, the two facilitators running the experiment trained extensively for this task and pre-trial recordings verified they were reliably able to do this with 98% accuracy.

6 Conclusion

The goal of the present work was to investigate whether it is possible to draw a persons' attention in a certain direction, regardless of the exact position the D-sounds was emitted from. Moreover, the respondents must have had no prior training, not draw on cultural references or previously learned sound patterns or be able to compare with other sounds.

The A and B experiments confirm this is possible to achieve for the up-down directions, using increasing or decreasing pitch patterns. Significant statistical results were obtained and close to 100% of the respondents stated they had experienced the effect for this case. Less clear results were achieved for the right-left direction, although half of the respondents replied they had associated the effect with the D-tones.

We conclude that sound patterns with a clear in- or decrease in pitch can indeed be used to direct an untrained users' attention up- or downwards. This is useful for a broad range of products and situations. In particular, for devices offering a number of spatially distinct "interaction points", such as ATMs, gasoline pumps and modern parking meters. These are often equipped with a small loudspeaker placed within the body for sound feedback. Here D-sounds can help to achieve a better user experience, lower production costs compared to implementing visual elements and most likely a simpler and more aesthetic design.

6.1 Further Work

Additional studies could be carried out to identify D-sounds with a perceived left-right directional effect, as this was not demonstrated fully in our experiment. Eye-tracking could be used to investigate in more detail, whether the responses are reflected or un-reflected. Finally, as this was a lab experiment, it would be useful to validate the results in a real-life situation: By designing D-sounds and include them in a usability trial with naïve users, measures such as task completion times and user satisfaction could reveal, if the desired effect is indeed present.

Acknowledgements. We received financial support for this work from Aalborg University and Innovation Fund Denmark, through the Danish Infinit network. Musician Keld Bauman composed and rendered the D-sounds for us. Bang and Olufsen A/S provided the loudspeakers and amplifiers used in the experiment. Finally, we thank all our respondents for lending us their ears and time.

References

1. Pratt, C.C.: The Spatial Character of High and Low Tones. Journal of Experimental Psychol. **13**, 278–285 (1930)
2. Blauert, B.: Spatial Hearing – The Psychophysics of Human Sound Localization. The MIT Press, Cambridge (1997). Revised edition
3. Roffler, S.K., Butler, R.A.: Factors that influence the localization of sound in the vertical plane. J. Acoust. Soc. Am. **43**(6), 1255–1259 (1968)
4. Mudd, S.A.: Spatial stereotypes of four dimensions of pure tone. J. Exp. Psychol. **66**(4), 347–352 (1963)
5. Simon, J.R., Mewaldt, S.P., Acosta Jr., E., Hu, J.-M.: Processing auditory information of two population stereotypes. J. Appl. Psychol. **61**(3), 354–358 (1976)
6. Walker, B.N., Ehrenstein, A.: Pitch and pitch change interact in auditory displays. J. Exp. Psychol.: Appl. **6**(1), 15–30 (2000)
7. Walker, R.: The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. Percept. Psychophys. **42**(5), 491–502 (1987)
8. Gaver, W.W.: Auditory icons: using sound in computer interfaces. Hum.-Comput. Interact. **2**, 167–177 (1986)
9. Gaver, W.W.: The SonicFinder, an interface that uses auditory icons. Hum. Comput. Interact. **4**, 67–94 (1989)
10. Blattner, M., Sumikawa, D., Greenberg, R.: Earcons and icons: their structure and common design principles. Hum.-Comput. Interaction **4**(1), 11–44 (1989)
11. Brewster, S.A., Wright, P.C., Edwards, A.D.N.: A detailed investigation into the effectiveness of earcons. In: Kramer, G. (ed.) Auditory Display, pp. 471–498. Addison-Wesley, New York (1994)
12. Gaver, W.W.: Auditory Interfaces. Handbook of human-computer interaction (1997)

http://www.springer.com/978-3-319-60476-3

Advances in Communication of Design
Proceedings of the AHFE 2017 International Conference on Human Factors in Communication of Design, July 17−21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA
Ho, A.G. (Ed.)
2018, X, 153 p. 55 illus., Softcover
ISBN: 978-3-319-60476-3