Utilizing Audio Cues to Raise Awareness and Entice Interaction on Public Displays

Hannu Kukka, Jorge Goncalves, Kai Wang, Tommi Puolamaa
Julien Louis, Mounib Mazouzi, Leire Roa Barco
Center for Ubiquitous Computing, University of Oulu, Finland
{firstname.lastname}@ee.oulu.fi

ABSTRACT
We present a study on the use of audio-based cues to help overcome the well-known issue of *display blindness*, *i.e.* to help people become aware of situated interactive public displays. We used three different types of auditory cues based on suggestions from literature, namely *spoken message*, *auditory icon*, and *random melody*, and also included a *no-audio* condition as control. The study ran for 8 days on a university campus using an *in-the-wild* design, during which both qualitative and quantitative data were gathered. Results show that audio in general is good at attracting attention to the displays, and *spoken message* in particular also helps people understand that the display in question is interactive.

Author Keywords
Public display; audio; display blindness; awareness; interaction; earcon; auditory icon; spoken message; melody.

ACM Classification Keywords
Human-centered computing~Human computer interaction (HCI); Human-centered computing~Empirical studies in HCI

INTRODUCTION
The use of auditory cues to convey information to users has been explored in the field of human-computer interaction since the late 1980's. These audio cues are sometimes referred to as *earcons* [1] or *auditory icons [8]*, and their overarching goal is to provide the user with intuitive, context-dependent information using sound as a channel, thus potentially reducing cognitive load imposed by dense visual displays of icons in menu structures. More recently, audio notifications have become commonplace as mobile devices have become pervasive [7]. Since the device is most often carried in a pocket or bag, the user is unable to see the screen as notifications arrive, and hence these devices often rely on a combination of tactile (vibration) and auditory (sound) feedback to alert the user of an incoming message. Simultaneously, interactive public displays are becoming a pervasive fixture in urban environments [22]. While often perceived as providing useful services (*e.g.* [13]), these displays frequently suffer from poor discoverability due to factors such as *display blindness* [16] or *interaction blindness* [22].

In this paper we present a study aimed at exploring the use of auditory cues to attract attention and entice interaction on situated interactive public displays. Based on recommendations from literature, we designed an in-the-wild experiment utilizing three types of auditory cues: *spoken message* (*e.g.* [4]), *auditory icon* (*e.g.* [8]) , and *random melody* (*e.g.* [1]). We also included a *no-audio* condition as a control. The study ran for 8 working days using 4 displays on a university campus, during which both qualitative and quantitative data were collected.

RELATED WORK
People rarely seek public displays actively, but rather encounter them in a serendipitous manner [10,18]. When encountering such a display, people need to first become aware of the display device itself. In a cluttered environment such as a city center this is not trivial. The tendency of people to overlook displays, also known as *display blindness*, has been identified in previous research [11]. Müller et al. [18] noted that many displays fail to attract enough attention of passers-by because they vanish in the clutter of things in public space that compete for attention. Huang and Borchers [11] found that using physical items nearby a display may help draw the attention of passers-by to the display itself, but only if the items in question come to the attention of a person before passing a display.

After noticing a display device, people need to become aware of its interactive affordances. This is known as *interaction blindness* [22], and refers to the fact that people do not realize that a display is touchable. Researchers have approached the problem of enticing interaction with public displays from various perspectives. Brignull and Rogers [2] identified three 'activity spaces' leading up to direct interaction activities, where people actively interact with a display. Ju et al. [12] studied so called 'attract loops' in enticing users to interact with an information kiosk. Kukka et al. [15] studied various atomic components of on-screen visual signals meant to attract attention and entice interaction on public displays.

DIS 2016, June 04 - 08, 2016, Brisbane, QLD, Australia
 ACM 978-1-4503-4031-1/16/06…$15.00
DOI: http://dx.doi.org/10.1145/2901790.2901856

While the use of audio seems like an intuitive way to raise awareness and potentially also communicate interactivity of situated public displays, to the best of our knowledge its use has not been systematically explored in previous work. However, previous research has employed many types of sound when designing interactive systems. While there are small inconsistencies in literature on the taxonomy of these sounds [1,6,7], the more prominent ones are: *spoken message*, *non-speech* and *random*. Spoken message presents the more direct and understandable approach in that users inherently know the meaning of the sounds, assuming it is in a language they understand. Previous work has highlighted how the human ear is focused primarily on distinguishing speech from all other sounds [20], making it more likely to be picked up. However, it can take longer to convey a spoken message than its non-speech counterparts, and a particular spoken message is not universally understood as only those familiar with the particular language will understand it [7].

On the other hand, non-speech stimuli can be easier to differentiate in speech based communication environments [6]. One type of non-speech stimuli are *earcons*, which consist of nonverbal audio messages used to provide information to the user about a computer object, operation or interaction [1]. A more recent definition states that all non-speech sounds that do not imitate everyday sounds can be considered an earcons [23]. The major weakness of earcons is their lack of semantic relationships to their referents [7]. This can be a limitation when designing earcon notifications, as they need to be learned "without benefit of past experience" [9]. Given the in-the-wild nature of public displays deployments, we opted to not use earcons as it would require training users to understand the meaning associated with them. Instead, we used another type of non-speech stimuli, an *auditory icon*. Auditory icons are sounds designed using the concept of 'everyday listening'. Auditory icons are conceptually more similar to graphical icons than earcons, as they utilize a metaphor that relates them to their referents [7]. Examples of auditory icons include 'shattering dishes' for dropping an object into the recycle bin [8] or 'door slamming' for remote users logging out of a network [3]. The main advantage of auditory icons is that because of the metaphors they utilize, training requirements can be kept to minimum, making them usable for in-the-wild settings. On the other hand, the major disadvantage of auditory icons is that virtual events and actions do not always lend themselves to everyday sound metaphors [7]. Furthermore, depending on the context of use, it is possible to confuse auditory icons with actual environmental sounds [3]. Finally, the *random* category consists of sounds that cannot be classified as any of the aforementioned auditory types [1].

STUDY

We followed a design similar to the study on the effect of atomic visual elements on enticing interaction on public displays by Kukka *et al.* [15]. For the purposes of the study, we deployed displays in populated areas of the university campus, *i.e.* along busy walkways and corridors (Figure 1 left). Spaces where people sit down and spend time such as cafeterias or restaurants were intentionally left out, as the use of continuous audio cues in such locations was deemed likely to become disturbing. The displays implement a scheduler that rotates each audio signal (and the no-audio condition) to counter-balance the effect of location so that each signal was shown on each display on a separate day, and no two signals were active simultaneously. However, since we used only 4 displays instead of 8 (as in Kukka *et al.*), our study ran for two "rounds" (8 working days in total), so that each signal was tested in each location twice.

The three audio cues used in the study are 1) *Spoken message*, 2) *Auditory icon*, and 3) *Random melody*. For the first cue, we opted to use a spoken message with the phrase *"This display is interactive"* in English, delivered by a professional female actress with a native (US) accent. Even though the study was conducted in a non-English speaking country, the international nature of a university campus warrants the use of English as a universal language. The actress was asked to deliver the line using a neutral tone without emotional tones or accent, as these can influence the effectiveness of the message [20]. For the second cue, we use an auditory icon of a female voice making an "*Ahem*" sound, as if politely asking for attention. The 'Ahem' sound has been shown to be an effective auditory icon to draw attention towards a device [7], and it was deemed distinct enough from other potential sounds in the environment. Finally, for *random melody*, we utilized a sequence of 3 random notes taken from the C major scale and played in a random order.

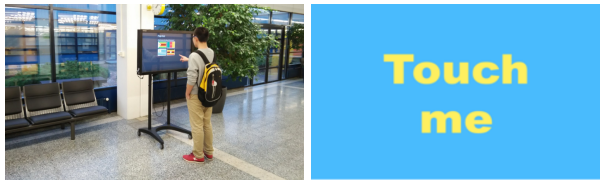

Figure 1. A user interacting with one of the displays during the deployment (left) and visual signal used in the study at maximum relative size to the screen (right).

We opted to use a female voice for both spoken message and auditory icon, as previous research has shown that women have more positive implicit associations with a female voice than with a male voice, whereas men either prefer a female voice (*e.g.* [5]) or are neutral towards one (*e.g.* [17,21,24]). Further, current practice with GPS navigation systems and audio-based digital assistants such as Siri and Cortana all favor the female voice, making it a natural choice for this experiment as well. All audio cues had the length of 1.8 seconds, and a volume level of ~55dB played back through the integrated speaker in the display frame. The audio cues are triggered when a camera placed on top of a display recognizes motion, *i.e.* someone passing by. We implemented a timeout in order to make the displays less disturbing by only playing the sound at 10

second intervals. The system also logs each trigger event, and this information can be used as proxy for the number of people passing by a given display on a given day. Also, again following Kukka *et al.*, we placed a visual signal on the screen in all conditions to investigate if an audio signal would provide an additional benefit over the visual signal alone. The signal is identical to the signal identified as most effective by Kukka *et al.* in their study, *i.e.* a colored and animated "Touch me" text (Figure 1 right). The visual signal contains yellow text on a blue background, and the text follows an anchored grow/shrink animation.

The displays follow a simple interaction model, where the first touch on the screen triggers a game where the user is asked to identify the flag of a certain country from four possible options. The game was implemented as a plausible reason for having the displays on campus without making it obvious that the study is about the use of audio cues. After the user completes the game or there are no touches for 30s, the display reverts back to showing the visual signal and the current audio cue. The system automatically logs all interactions performed on the displays, including the ID of the display, the currently active condition, and the timestamp of the first touch to the screen. Further data was collected through unobtrusive observations on two different displays (32 hours), and structured in-situ interviews (N=48) conducted with members of the public after they had interacted with a display. To reduce the likelihood of bias, we approached people after they had left the display in order not to obstruct the display, to prevent them being able to hear the sounds, and to ensure that passers-by could not infer that people using the display were being interviewed.

RESULTS

Log Data

In total, the displays attracted 1418 touches over the 8 days of deployment (M=177.25, SD= 62.01). The most popular condition was *spoken message* (N=410), followed by *auditory icon* (N=397), *random melody* (N=379) and finally *no-audio* (N=232). On average, the audio cues were triggered 2170 times per day of deployment and per deployment location. Figure 2 shows the average number of touches per day for each of the conditions.

These numbers were then normalized based on the total number of triggers for each day of deployment before conducting any statistical analysis. A Kruskal-Wallis H test showed that there was a statistically significant difference in number of touches between the different conditions, $\chi2(3) = 9.141$, $p = 0.03$, with a mean rank number of touches of 23.19 for *spoken message*, 16.56 for *auditory icon*, 17.19 for *random melody* and 9.06 for *no-audio*. Pairwise comparisons between the different conditions only showed statistically significant difference between *spoken message* and *no-audio* ($p = 0.02$). These results suggest that audio in general, and *spoken message* in particular, were more successful in enticing interaction than solely relying on a visual signal.

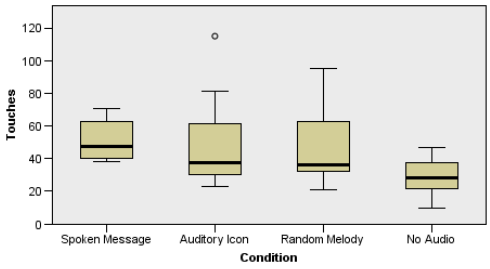

Figure 2. Average number of touches per day for each condition.

In-situ Observations and Interviews

We conducted 32 hours of in-situ observations (2 hours/day on two different displays). During this period, we observed a total of 752 interactions. As expected, audio is very good at attracting attention and helping passers-by notice the displays. People often passed a display without apparently noticing it, but when the audio cue triggered they turned back, sometimes several times, to identify the source of the sound (also known as the *landing effect* [19]). Often, the visual 'touch me' text on the display then appeared to persuade them to approach and begin interaction. People in groups exhibited stronger reactions than those who were alone, with groups stopping to engage in playful behaviors such as imitating the 'Ahem' sound, laughing and talking about the sound, intentionally moving in front of the camera to trigger the sound, or competing to see who can touch the display first. We also observed people passing the displays, immersed in an activity such as using their mobile phone and hence not paying attention to their surroundings. Without the audio cue, these people would have completely missed the display, but the cue caught their attention and made them shift their focus away from the phone and towards the source of the sound. Some participants also attempted to identify to source of the sound more accurately by walking around the display to see the speakers (which are built into the frame and hence not visible), or pressing their ear to the display frame.

Furthermore, we conducted 48 (21 male, 27 female) in-situ interviews at different displays. The interviews were structured around the following questions:

1. You just interacted with one of the public displays here on campus. What made you stop and interact with the display?
2. Do you remember hearing a sound coming from the display?
    a. Can you describe the sound for me?
    b. Did you think the sound was disturbing?
    c. Was it clear to you that the sound was coming from the display?
    d. Did the sound tell you that the display in question was interactive?
3. Do you remember what was shown on the display?
    a. Can you describe what was shown for me?
    b. Did the text tell you that the display was interactive?

4. In your opinion, what is better for drawing your attention to a public display, sound or text?
5. In your opinion, which is better for letting you know a display is interactive, sound or text?
6. In your opinion, which is better for attracting you to actually stop and interact with a display, sound or text?

Regarding question 1, 44% of the respondents identified the textual cue as their main reason for stopping to interact, while 31% stated that the audio cue was their main motivation for interacting. The remaining 25% reported interacting due to curiosity and not because of either cue. Participants were good at remembering the audio cues (Q2), with 90% of interviewees remembering the auditory icon, 75% remembering the spoken message, and 67% remembering the random melody. Spoken message was deemed good at communicating interactivity (88%). However, the other two audio-based cues did not fare as well, with 38% of respondents in the auditory icon condition and 0% of respondents in the random melody condition reporting that these cues were good at communicating interactivity.

All participants reported remembering the visual signal shown on the screen (Q3), and thought that the text was good at communicating interactivity. 51% of respondents identified audio as being better at drawing their attention to a display, with 37% identifying the visual cue as being better and 12% stating that the combination of both works best (Q4). Further, regarding question 5, 64% of respondents thought that the visual cue is better at communicating interactivity, with 18% claiming sound is better and 18% identifying the combination of both as the most efficient. Finally, regarding question 6, 55% of respondents reported the visual cue as being better for attracting them to stop and interact with a display, with only 18% identifying sound as being better and 27% naming the combination of both as best.

DISCUSSION AND CONCLUSION

The presented study clearly illustrates the benefits of utilizing audio-based cues in attracting attention towards public displays: all conditions with audio clearly outperformed the no-audio control condition. However, interview data showed that participants identified the visual signal as their main motivation for stopping to interact (Q6). Therefore, we conclude that while visual is good for communicating interactivity, audio is better at capturing attention (Q4). The *spoken message* cue also proved to work well in communicating interactivity.

However, given the implicitly *public* context of interactive public displays, the use of audio entails more complex underpinnings and implications. The first issue to consider is the variance of people in a given space. In an environment with low variance such as a workplace or local coffee shop, people will likely learn a display's interactive affordances after the first few sessions with the device; after that, continuous audio messages are likely to become disturbing. Hence, in locations where the same people can be expected to hear the audio over and over, it is important to control the potential disturbance by extending the time between audio notifications.

Secondly, some locations are simply not well suited for audio-based notifications to begin with: for instance, a library would likely be unsuitable, whereas an already busy environment such as a shopping mall or train station would likely benefit from utilizing audio to attract people's attention towards *e.g.* information displays. Further, when considering using audio as an attractor to a public display, it is important to first consider the ambient noise level and the suitability of added audio in the given space. An interesting future direction is to include a microphone in the display frame to detect ambient noise level, and adjust the audio cues volume accordingly within a reasonable threshold. This would again help lessen the potential disturbance caused by the use of audio cues, as volume could always be kept at an appropriate level instead of designing for a "loudest case" scenario.

Finally, the audio cue itself should be carefully considered: even though all the cues used in this study proved to work well in attracting attention, only the spoken message was found to be effective at communicating interactivity. Even though auditory icons and earcons surround us every day (audio-based notifications for email and SMS, alerts for meetings, ringtones of mobile phones…), they necessarily rely on the user either setting his/her desired notification sounds and thus knowing them [7], or using universally recognizable sounds. In busy urban settings the auditory landscape is rather saturated already: cars passing by, traffic lights giving off auditory information for accessibility reasons, people chatting, mobile phones ringing, and so forth [14]. In such settings an auditory cue needs to be very explicit and clear, and people should be able to intuitively identify its source and purpose. For this reason, a spoken message clearly identifying the source ("this display"), and purpose ("is interactive") can be considered as optimal. Of course, language can then present an issue: we opted to use English as the deployment setting was a university campus where people can be expected to know English. However, in other settings this may not be the case and a message in one language may not be comprehensible to everyone. In such settings, the use of multi-lingual messages or an easily understandable auditory icon such as the one used in this study should be considered. In less complex settings, however, other types of auditory cues are likely to work equally well, especially when coupled with a visual signal on the screen identifying interactive affordances. In our study, users predominantly identified audio as being good at capturing attention, especially when the user is occupied with some activity (*e.g.* looking at their phones while walking or talking with friends). This is an important finding, as making people aware of a display (overcoming display/interaction blindness) is the crucial first step in any public display interaction.

REFERENCES

1. Meera M. Blattner, Denise A. Sumikawa and Robert M. Greenberg. 1989. Earcons and icons: Their structure and common design principles. *Human-Computer Interaction* 4, 1: 11-44.
2. Harry Brignull and Yvonne Rogers. 2003. Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT*, 3, 17-24.
3. Jonathan Cohen. 1994. Monitoring background activities. In *Santa Fe Institute Studies In The Sciences Of Complexity, vol.* 18, 499-499.
4. Tilman Dingler, Jeffrey Lindsay and Bruce N. Walker. 2008. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. *Proceedings of the 14th International Conference on Auditory Display (ICAD2008)*
5. Alice H. Eagly, Antonio Mladinic and Stacey Otto. 1994. Cognitive and affective bases of attitudes toward social groups and social policies. *Journal of Experimental Social Psychology* 30, 2: 113-137.
6. Judy Edworthy and Neville Stanton. 1995. A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology. *Ergonomics* 38, 11: 2262-2280.
7. Stavros Garzonis, Chris Bevan and Eamonn O'Neill. 2008. Mobile service audio notifications: intuitive semantics and noises. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, 156-163.
8. William W. Gaver. 1986. Auditory icons: Using sound in computer interfaces. *Human-computer interaction* 2, 2: 167-177.
9. William W. Gaver. 1997. Auditory interfaces. *Handbook of human-computer interaction* 1: 1003-1041.
10. Jorge Goncalves, Simo Hosio, Yong Liu and Vassilis Kostakos. 2014. Eliciting situated feedback: A comparison of paper, web forms and public displays. *Displays* 35, 1: 27-37.
11. Elaine M. Huang, Anna Koster and Jan Borchers. 2008. Overcoming assumptions and uncovering practices: When does the public really look at public displays? *Pervasive Computing,* 228-243.
12. Wendy Ju and David Sirkin. 2010. Animate objects: How physical motion encourages public interaction *Persuasive Technology,* 90-101.
13. Hannu Kukka, Vassilis Kostakos, Timo Ojala, Johanna Ylipulli, Tiina Suopajärvi, Marko Jurmu and Simo Hosio. 2013. This is not classified: everyday information seeking and encountering in smart urban spaces. *Personal and Ubiquitous Computing* 17, 1: 15-27.
14. Hannu Kukka, Anna Luusua, Johanna Ylipulli, Tiina Suopajärvi, Vassilis Kostakos and Timo Ojala. 2013. From cyberpunk to calm urban computing: Exploring the role of technology in the future cityscape. *Technological Forecasting and Social Change, 84,* 29-42.
15. Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge Goncalves and Timo Ojala. 2013. What makes you click: exploring visual signals to entice interaction on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1699-1708.
16. Nemanja Memarovic, Sarah Clinch and Florian Alt. 2015. Understanding display blindness in future display deployments. In *Proceedings of the 4th International Symposium on Pervasive Displays*, 7-14.
17. Wade J. Mitchell, Chin-Chang Ho, Himalaya Patel and Karl F. MacDorman. 2011. Does social desirability bias favor humans? Explicit--implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior* 27, 1: 402-412.
18. Jörg Müller, Florian Alt, Daniel Michelis and Albrecht Schmidt. 2010. Requirements and design space for interactive public displays. In *Proceedings of the international conference on Multimedia*, 1285-1294.
19. Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt and Florian Alt. 2012. Looking glass: a field study on noticing interactivity of a shop window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 297-306.
20. Clifford Nass and Li Gong. 2000. Speech interfaces from an evolutionary perspective. *Communications of the ACM* 43, 9: 36-43.
21. Brian A. Nosek and Mahzarin R. Banaji. 2001. The go/no-go association task. *Social cognition* 19, 6: 625-666.
22. Timo Ojala, Vassilis Kostakos, Hannu Kukka, Tommi Heikkinen, Tomas Linden, Marko Jurmu, Simo Hosio, Fabio Kruger and Daniele Zanni. 2012. Multipurpose interactive public displays in the wild: Three years later. *Computer* 45, 5: 42-49.
23. Antti Pirhonen, E. Murphy, G. McAllister and W. Yu. 2006. Non-speech sounds as elements of a use scenario: a semiotic perspective. *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*
24. Laurie A. Rudman and Stephanie A. Goodwin. 2004. Gender differences in automatic in-group bias: why do women like women more than men like men? *J Pers Soc Psychol* 87, 4: 494-509.