AUDITORY NOTIFICATION OF CUSTOMER ACTIONS IN A VIRTUAL RETAIL ENVIRONMENT: SOUND DESIGN, AWARENESS AND ATTENTION

Kjetil Falkenberg, Emma Frid
KTH Royal Institute of Technology
Stockholm, Sweden
kjetil@kth.se, emmafrid@kth.se

Martin Ljungdahl Eriksson
University West
Trollhättan, Sweden
martin@soundmark.se

Tobias Otterbring
University of Agder
Kristiansand, Norway
tobias.otterbring@uia.no

Sven-Olov Daunfeldt
Institute of Retail Economics
Stockholm, Sweden
sven-olov.daunfeldt@huiresearch.se

ABSTRACT

In this paper, we introduce sonification as a less intrusive method for preventing shoplifting. Music and audible alerts are common in retail, and auditory monitoring of a store can aid clerks and reduce losses. Despite these potential advantages, sonification of interaction with goods in retail is an undeveloped field. We conducted an experiment focusing on peripheral auditory notifications in a virtual retail environment, evaluating aspects such as awareness and attention, sound design and noticeability, and localization of event sounds. Results highlighted behavioral differences depending on whether users were informed about the presence of auditory notification sounds or not. The alerts did not cause distraction or annoyance and we suggest that the findings give a promising starting point for future studies and investigations focused on improving the auditory environments in retail.

1. INTRODUCTION

Retailers state that shoplifting has a significant negative impact on profitability [1], and in the United States alone the total loss due to shoplifting has reached US$10 billion yearly [2]. According to [1] the most effective countermeasures against shoplifting involve human factors, and in particular security guards, which increase costs significantly. However, video or security guard surveillance can be perceived as an invasion of privacy. In this paper, we introduce sonification as a less intrusive method for surveillance. The research focus is on reactions to notification sounds in order to investigate aspects such as attention, awareness, sound design and noticeability through analysis of localizing sound sources.

Auditory notifications in stores are meant to attract attention from store employees and shoplifters, but without disturbing regular customers. There are thus several different types of visitors in a store to consider: 1) the store employee who manages customers and sales, and who also keep watch of the premises; 2) the shoplifters, who can be classified according to their intention, see e.g. [3]; and 3) the regular purchasing or browsing customers. In the present study, we consider two groups: those aware of a surveillance system (employees, shoplifters, possibly some purchasing customers), and those unaware (most regular customers), represented by participants with or without knowledge of notification sounds, respectively.

The auditory system is dependent on attention and built to process simultaneous and overlapping stimuli [4]. Auditory monitoring can free up cognitive resources; e.g., people working in security operations centers were aided by sonification in ways that enabled peripheral monitoring [5]. However, few studies have tested sonification in retail and its impact on key outcomes, such as the consumers' and employees' attentional mechanisms, the former of which is a key factor in predicting consumer choice [6, 7].

The implementation of sound alerts constitutes a less intrusive way of ensuring safety and avoiding shoplifting in the retail environment compared to existing solutions, such as monitoring through video cameras. Thus, the current research may have privacy implications through innovative sound strategies, whereby retail stores could potentially replace or reduce video monitoring in favor of less intrusive sound alert solutions.

From the above, we believe that sonification can be a good candidate for designing a system for improved store surveillance and shopping experience, and a novel approach in the field of auditory displays. In the current study we delimit sonification to systematically connecting avatar actions in VR to static sound recordings. In particular, we use an event-based approach for monitoring state in a multimodal environment [8].

1.1. Aim

The purpose of this study is to investigate the following research question: *Do subjects assigned a store clerk role, i.e., who know that they will be exposed to notification sounds, react more to such sounds than subjects assigned a browsing customer role, i.e., who are unaware of notification sounds?* In particular, we sought out to investigate how these two conditions affected the number of noticed sound events measured by head movements in direction of the sound source, and how important sound design in terms of having congruent versus incongruent sounds is in this context. Our assumptions are that the “knowing” subjects react more to the sonifications than the “unknowing”, and that the “unknowing” will be more disturbed by the incongruent sounds.

1.2. Previous results

Our design for the sonifications were validated in [9]. There, we also describe the simulated store environment in VR, with participants being represented through an avatar. In [9] we propose a first set of design guidelines for short alert sounds in a retail environment. Alert sounds can:

- be congruent with and contextually fit the environment of where they are played,
- be played at a relatively low volume, preferably just below background music,
- be short in length, around one second,
- be designed without much attention towards attack sharpness,
- be used without concerns of growing sensitivity over time,
- be incongruent, if designed with care.

2. BACKGROUND

The focus of the current research is on peripheral auditory notifications in a (virtual) retail environment. Several previous studies have focused on peripheral notifications in which audio cues are embedded in background music (see e.g. [10, 11, 12]). Some work has also focused specifically on notifications in the retail environment. For example, in [13], personalized ambient soundscapes allowing for notification services through non-speech audio cues were embedded differently in background music depending on the event and current position of the user. In [14], authors introduce a method to make personalized music freely selectable for each customer in a retail environment. Their system allows customers to create an electronic shopping list where each item is associated with recommended music. When the customer enters a specific product department, the music track associated with the product is played. The system also notified customers by playing non-speech sounds mixed into the background when you approached products similar to those listed in the shopping list.

The perceptual and cognitive knowledge of an auditory environment is grounded in our ability to learn from previous experiences. We use this knowledge in relation to a context to predict which sounds that are likely to appear [15], but also to reject interpretations incongruous with a context [16]. When audiovisual sensory information is unrelated it leads to an uncertainty of interpretation, causing an attentional focus on identifying what is incongruent [17].For instance [18] showed that sounds incongruent with the background sound environment are easier to identify and detect than contextually congruent sounds.

Perceptual attention is usually defined in terms of internal processes that help us extract relevant information from a complex environment [19]. The well-known *cocktail party effect* highlights the ability of listeners to move their attention around a busy environment [20, 21], and has been seen as an example of listeners ability to derive information from a stream of speech sounds in the presence of other sound streams [19]. Indeed, information can be processed even if not in the foreground of attention [22, 23], meaning that the allocation of attentional resources depends on a variety of factors, such as personal interests, prior experiences, and expectations (top-down factors), but also the cognitive capacity of a person at a given point in time, and the specific social settings in which the person performs an action [24, 25, 26].

A detailed understanding of the way end users inhabit their environments, as well as how they interact with the existing auditory setting, is required in order to design auditory interfaces that integrate effectively in such environments [27]. An important term in this context is the notion of *awareness*. In the context of ambient displays; i.e., displays which present information through subtle changes in light, sound, and movement; awareness can be defined as *“(...) the state of knowing about the environment in which you exist; about your surroundings, and the presence and activities of others”* [28]. Peripheral awareness of audio has, for example, been discussed in the phenomenological study of the auditory environment in a chemical factory presented in [29]. In this work, the author discusses the paradox of perceptiveness versus unobtrusiveness of sound events; i.e. the challenge of enabling sound events to fade into background of awareness if not needed but still be perceptive enough to inform a user about a change, introducing the notion of *smooth notification*.

The work presented in this paper relies on measuring *sound localization* through head movements in a VR space. A pioneering researcher in the field of sound localization was Lord Rayleigh, who presented work on the localization of pure tones in the lateral dimension, commonly referred to as the *duplex theory* of sound localization. Rayleigh concluded that subjects could not easily discriminate between front versus back locations of pure-tone stimuli [30], although this was possible for sounds with broader bandwidths [31]. More recent work has suggested that the the duplex theory is somewhat incomplete, and that physical cues other than interaural time and level differences are used in sound localization, such as pinna filtering [32]. Moreover, front-back error tend to be reduced by allowing for head rotation [33, 34]. It has been suggested that head movement is part of the listening experience since it allows for sensing the spatial distribution of parameters [35]. An important concept in this context is the head-related transfer function (HRTF), i.e., the direction-dependent acoustical transfer function from a sound source to a listener’s eardrum [36]. An overview of basic principles and applications of head-related transfer functions and virtual auditory display is presented in [37].

The relation between head movements in VR experiments focused on sound has been explored in numerous previous studies (see e.g. [35, 38, 39]). In the current work we focus on auditory-localisation of events in a VR environment based on head pointing, meaning that the participant turns the body towards a perceived sound source and point its direction with the head. It is well known that the method used for collecting judgements in auditory-localisation experiments strongly influences the accuracy of subject’s responses [40]. Egocentric pointing methods, in which the participant’s body (including the head) is directed towards the sound source, have been shown to be more precise then exocentric methods, in which the participants report perceived sound localisation on a 2D/3D graphical interface [41, 42, 43]. Head pointing has shown many advantages over other methods when applied in previous sound source localisation studies, despite issues related to localising highly elevated targets [43, 44, 45].

When it comes to sound design, the need to design more aesthetically pleasing sonification and alert sounds has been stressed by several scholars [46, 47]. The importance of designing alert sounds with a high level of ecologic validity, here conceptualized as a match in terms of function, has also been widely discussed [48, 49]. Ecological validity conceptualized as a match in terms of good recordings or realistic sound simulations are not necessarily the most efficient design strategy to convey information, as shown in research literature on sound objects and cartoonification [50]. Instead, low-level models based on simplifications of the target simulation, for instance by exaggerating certain acoustic aspects, have proven to be very effective, and contextual attachment should thus be an important design parameter.

3. METHOD

A total of 16 participants (9F, 7M, age 24–53 years) took part in the experiment. The experiment is designed as follows: we track a participant's head movements in a virtual store and match the head direction to sounds initiated by avatars interacting with items in the shop. We consider each sound to be a discrete sonification of removing an item from a shelf. In the following section, we describe the VR and sound environments, and the data collection.

3.1. Virtual environment

Instead of conducting the experiment in a real store, we built a virtual clothing store for Oculus Quest head-mounted display (HMD) in the Unity 3D game engine to simulate a shopping experience. The kind of HMD VR used in this study is defined by providing 3D stereo vision, surround vision and user dynamic control of viewpoint [51]; in addition, the HMD is designed so that most over-ear headphones can be worn comfortably to play spatialized sound. These features, when implemented together, provide for an immersive experience where the user is perceptually shielded from the surroundings, but where the experience matches a real world. Studies have showed that the sense of presence and immersion is generally high for HMD [52].

In the implementation, participants used Audio-Technica ATH-M50X headphones. The chosen spatialization mode in the software was without corrections for vertical head displacement, i.e., only based on horizontal movements.

Figure 1: The virtual store environment. Left image: The floor layout with the six locations for alert sound sources marked with red circles. Right image: A design-stage screenshot from VR with the avatars that will act as customers and clerks.

The defined store was spacious, with 700 square meters consisting of three rooms, see Fig. 1. The space consisted of a main shopping area with shelves, tables and counters, a fitting room section, and a smaller room approached through an open door, with some shelves and a table. One adjacent room and a store entrance were also included in the model, but not used for the experiment. The tables and shelves contained garments and accessories, for the most shirts, pants, hats, backpacks, belts, and handbags. Six preprogrammed avatars (four female and two male characters) who acted as customers would move around and approach the store goods to look at the different items (but not directly grab them).

3.2. Sound environment

A generic deep house background music track was playing in the store continuously. Although music was played from a set of loudspeakers placed in the ceiling, the experience in the VR environment was that the music was ever-present and at a constant sound level. In addition to and mixed in with the music track, we played a soundtrack of ambient store sounds with cashier noises, entry chimes, mumbling, shoe scraping, among others; the recording was selected to match and add multimodal realism to the depicted scenery, without the ambition of recreating an exact sonic representation of a physical store.

When an avatar approached an item on a shelf, a sound alert was played from that position. While the alert sounds were spatially separated, sounds were not acoustically affected by walls and other objects, thus, the room acoustics resembled an open space. Six different locations, each with its own sound, were included, indicated in Fig. 1. We programmed in total 25 interactions with varying distances from the track of the participant's moving avatar, with temporal intervals between sounds ranging from 11 to 35 seconds. The sounds were played only slightly louder (a few dB) than the background music (see [9]). Each of the six sounds was played between 3–5 times. In two occurrences, the same alert sound was repeated, i.e., played from the same location twice in a row (but with a similar time interval as all other events).

The six different sounds all had a duration of 0.5–2.5 seconds, grouped by being either congruent or incongruent with the store settings. The congruent sounds resembled removing a piece of clothing from a hanger; two sounds were actual recordings of a clothing hanger, and one sound was a cartoonified, sweep-like sound. The incongruent sounds were distinctly different: a bird song, a time-stretched sweep sound, and wind chimes, all selected on grounds of their contextual detachment from the store environment. All sounds, including the background soundscape, are available for listening online.[1]

3.3. Experiment design

In this study, we were mainly interested in reactions to sounds triggered by avatars that move around the participant. As such, we only considered the participant's head movements directly connected to alert sounds. To create a comparable experience for the participants, the avatars' movements and actions were identical between trials. Moreover, the participant could not decide where to go, but was led around along a path in the store, to simulate a person (either observing store clerk or browsing customer, based on the instructions) walking casually between shelves and rooms. To facilitate a more natural experience of the fixed movement track and reduce the risk of motion sickness, the participant was encouraged to stand between and hold onto the back of two chairs while

[1]http://annexes.smcresearch.se/2021-ICAD-KF (a file with the movement data are also available at the link)

walking and turning on the same spot, actively following the virtual representation of the participant. Head movements, also those from turning the torso or the whole body, were unrestricted, and we recorded this horizontal rotation for further analysis.

We assigned participants randomly into one of two groups: "knowing" and "unknowing". Both groups were briefed that they would play a store clerk in an informal game or experience where you could not win or loose, and where the main objective was to look around and experience the VR store. The knowing group were given instructions that sound alerts or notifications can be triggered by the other avatars, and a task to, if they wish, look out for such events. The unknowing group were initially given ambiguous information that we tested machine learning algorithms for avatar interaction, but nothing about the sound environment.

3.4. Data collection and analysis

The test in the VR environment lasted for 10 uninterrupted minutes. The participants were briefly interviewed after the VR session. Head movements in terms of absolute position in the room and the horizontal rotation were sampled at 10 Hz and saved in the Oculus Quest HMD. The motion path of all the avatars, including the participant's avatar, were identical between sessions. In addition, the positions and timestamps of the played alert sounds were identical.

In the context of this work, we construed an event called "hit", i.e., an immediate and localized reaction to respective sound stimulus, when the participant's head orientation matches a sound source's location directly following its occurrence. The detection of hits was performed using an "inspection window" which was defined as the duration from the sound onset to a given offset. From experimentation, we set the parameters for angular range and inspection window manually. Then, we define the term "hit rate" as the ratio between the total number of hits (i.e., events within both the temporal window and the angular limit deviation from the sound's origin) and the total number of sound events.

The temporal window parameter defined the duration starting from the sound onset in which to consider reactions. Experimentation showed that the effect of changing this temporal window size was small (e.g., a 50% window size increase only resulted in an 8% hit rate increase). Furthermore, we observed a ceiling effect starting from less than four seconds after offset. We settled for an inspection window size from the sound onset until three seconds following its offset as default for all sound events. The angular range parameter defined the head orientation relative to the direction of the sound. Experimentation showed that widening or narrowing the hit angle ranges had little effect on the number of hits. We settled on $\pm$ 15 degrees from the direct line as a default deviation.

Movement data from the HMD sensors was collected to analyze differences between hit rates of unknowing and knowing subjects. We used t-tests and Chi-square tests with 5% significance level. The movement data is made available online at the link provided above for sound examples.

While our sample size is small in terms of the absolute number of participants, it is worth mentioning that each participant contributed with multiple data points, given the repeated exposure to auditory notifications and the additional data points collected between these sound alerts. Thus, despite the small sample size, comparisons between the two experimental conditions were based on hundreds of observations per condition, making the study sufficiently well-powered.

4. RESULT

In the following section, we present effects of awareness and attention, sound design and noticeability, and of the tendency to localize the sounds (hit rate). By "awareness and attention", we consider a perspective of active listening, while by "sound design and noticeability", the perspective is that of the alert sound being noticed in passive listening.

The result section include quotes from the interviews conducted following each VR session. We do not cover the sound design aspect which is, although critical, out of scope in this paper; this has been reported separately in [9].

4.1. General observations

We could measure reactions to sounds from the head movements, and that the two groups differed. As expected, the knowing had a significantly higher head movement activity than the unknowing, both in terms of velocity and in angular displacement (t-test, $p < 0.001$), see Fig. 2, which shows velocity measured as the derivative of movement. We also found a significantly higher activity during sound events for the knowing group compared to unknowing group (t-test, $p < 0.001$; all differences had a high effect size, Cohen's-d > 1.4). The knowing participants made almost double the amount of head movements during an sound event compared to the time in-between sound events, while the unknowing participants' head movements were similar for both conditions.

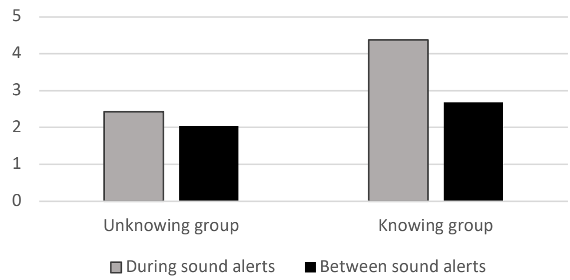

Figure 2: Head rotation velocity (measured in angular displacement per time unit) for both unknowing and knowing groups. The plot shows mean velocity during and between sound events.

Most participants had little to no previous experience with VR environments, but there is no evidence in the interview data that technology or the VR experience was problematic, nor that any suffered from motion sickness. When asked if they noticed anything out of the ordinary, all in the knowing group mentioned issues related to sounds, while all in the unknowing group answered either "no", or mentioned issues unrelated to sound (such as the variety of clothes).

4.2. Awareness and attention

The knowing participants outperformed the unknowing group significantly in terms of hit rate (χ^2, $p = 0.000$). On average, the unknowing reacted by looking in the direction of the sound source for 30% of all sound events, while the knowing reacted to 67% of the sound events. All except one participant in the knowing group performed better than the unknowing.

In the interview question about anything catching their attention, one participant in the unknowing group mentioned "*there is someone who is perhaps doing something shady*". In the knowing group, all except two connected sound events to actions by avatars, e.g., "*After a while I noticed that sound was connected to movements*", and also about shoplifting: "*[. . .] I heard noises and looked back and started to think if someone snatched. There were a few times I thought it was suspicious action.*", and customer behaviour: "*It's like a pattern. I felt that at some places there were specific sounds [. . .] the customers seemed like they needed help*".

It is possible that the overall attention dropped over time. Both groups showed a quick drop in hits after the initial couple of events, with a stable and slight decline following. Fitted, logarithmic regression curves were, however, inconclusive (with R-squared values of 0.09 and 0.19, respectively), see Fig. 3. A regression analysis on exposure shows difference between the groups ($p < 0.001$), but the short test duration contravenes making predictions about wearing effects and adaptation.

4.3. Sound design and noticeability

The duration of the sound alerts did not have a significant effect on hit rate for the knowing group. However, we observed a within-group difference for the unknowing group (χ^2, $p = 0.025$), where sounds longer than 2500 ms resulted in almost twice as many hits on average compared to sounds below 1500 ms. We found no significant differences for knowing versus unknowing between having a fast or slow onset time, with sound amplitude peaks at 100-300 ms and 700-800 ms, respectively.

Two sounds were designed to be particularly easy to notice, namely, a recording of a chirping bird (2500 ms, early onset), and a fast ascending stroke on metallic wind chimes (1500 ms, late onset). Indeed, the bird type sound was the most noticed among the unknowing group (χ^2, $p = 0.025$), while the very characteristic wind chimes got the opposite effect of the anticipated one, and it was the least noticed sound, see Fig. 4.

Three of the six sounds had intentional high congruence with the store context, being sounds inspired by clothing shop interactions. These were thus coupled to the experience and, assumable, more subtle and harder to notice through the ambience. We did not find any significant difference for either knowing or unknowing group between hit rates for either congruent or incongruent sounds. When comparing differences between means for hit rates, we observe that congruent sounds have on average 11% higher rate than incongruent sounds, where the knowing appears to be the most observant of congruent sounds, however, these results are inconclusive.

One interview question focused on the perceived sound level. Although the sound level was the same for all, half of the participants in the unknowing group mentioned that the sound level was too low, while all in the knowing group said it was good.

The unknowing group was specifically asked if they heard any sounds apart from the music. Two had noticed the bird song, one said "*perhaps some 'swish' sound*", but the rest had only noticed the entrance chime from the ambience soundscape. The knowing group was instead asked if sounds were clearly distinguishable. One participant in this group said "*there were not very many sounds I heard, maybe just two types and they did not come so often*", yet, this individual had a 72% hit rate. Another, with a hit rate of 76%, listed the following identified sounds in the interview: "*Bird sound, metallic drum, 'ritsch', telephone signal, glittering ring–chimes, dull dark sounds that felt like it's not how clothes sound, natural and unnatural sounds in addition to the music.*"

4.4. Localization of sound events

There was no measured effect on hit rate of neither distance between the participant's avatar and the sound source, nor the sound source placement in the store (χ^2-tests). Taking into account the head rotation velocity during sound alerts, which is high for the knowing, and hit rate for the same group, which is in average above 67% (see Figs. 2 and 4), the results indicate that knowing participants will predominantly turn and look directly towards the sound source and not in other directions.

One of the knowing responded in the interview that it was hard to localize sounds, although the hit rate for this participant was high, 68%. The participant who performed best, with 92% hit rate, said it was "*Pretty easy. But it was easier when you got used to the sounds.*". Two other participants, who also described the task as being pretty easy, only had 52% versus 55% hit rates, respectively. We did observe an increase in hit rate from the two instances of repeated sounds for the knowing group (χ^2, $p = .005$), but the result is inconclusive because of the limited data available for this comparison.

5. DISCUSSION

Analysis of collected data suggests that the sound alerts used in the current study were successful in guiding the attention of participants to the location of respective sonic event, but that there was a considerable difference between groups, both for hit rate and head velocity measures. For example, the unknowing group looked in the direction of the sound source for 30% of the events, whereas the knowing group reacted to 67%. Moreover, we observed a non-significant difference in hit rates for incongruent versus congruent sounds, with the unknowing group being less susceptible to congruent sounds.

The effect of sound duration on hit rate also appeared to be different in the two groups: for the unknowing group, sounds longer than 2500 ms resulted in almost twice as many hits on average compared to sounds below 1500 ms. Moreover, some interesting findings were observed from the interviews regarding sound level: although the sound level was the same for all participants, half of those in the unknowing group mentioned that the sound level was too low, while all in the knowing mentioned that is was good. It thus seems feasible to adjust the perceptiveness of sound alerts to achieve a good balance between noticing and overlooking them (shown in [53]), and that the alerts can be designed for different circumstances and needs (suggested in [47]).

The method used in auditory-localisation experiments strongly influences accuracy of subject's responses [40], and it is possible that other localisation methods than head movement tracking could be suitable in this context (especially if elevated targets are present). More realistic interactions, e.g. by allowing participants to touch objects with their hands, could also promote more accurate head movements (see e.g. [39]). In a real-world store setting, other means of determining reactions to sounds apart from head movement should be considered, for example measures of gaze. Such experiments on-site would also introduce a number of confounding variables and require additional considerations

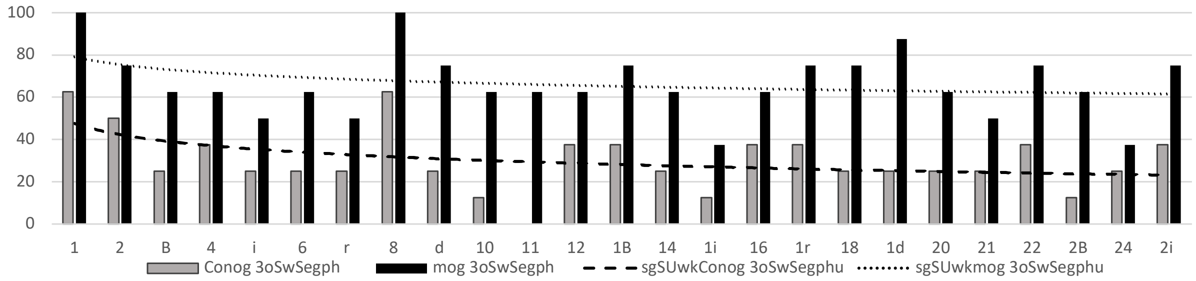

Figure 3: Hit rates for each sound event divided by participant groups (unknowing versus knowing), with fitted logarithmic regression curves (dotted: knowing $R^2 = .09$ and dashed: unknowing $R^2 = .19$).

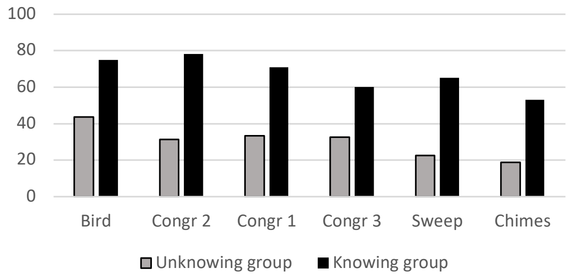

Figure 4: Hit rates for each sound ordered from highest to lowest average rate, divided by participant group (unknowing versus knowing). Three sounds are congruent, while three sounds (sweep, chimes, bird) are incongruent.

when it comes to ethics and privacy, which is the main reason why a VR environment was used in the current study.

When it comes to reactions to the sound design of the stimuli used in the current study, we obtained some unexpected results. One striking example is the low hit rate of the wind chimes sound, which we expected to be the most easy sound to notice. Why did it go more unnoticed compared to other sounds? Perhaps the perception of this sound was confused with the door entry bell from the ambient background, or the expectancy of such a sound or the sound of a cash register. Further testing is required to increase the level of detail in the design guidelines described in [9].

5.1. Limitations

The current work is not without limitations. It should be noted that the design, and the corresponding study-setup, has relatively basic research character. We are aware of the many limitations of VR in marketing research (e.g. [54]) and used this technology only to collect data related to perception, not consumer experience. While we document a set of behavioral differences as a function of whether users are (vs. are not) informed about the presence of auditory notification sounds, we have yet to test whether the mere awareness of such sounds may influence store employees' effort, ease, and efficacy in detecting customer theft, and whether alerts may create spillover effects on employees' satisfaction levels.

However, we demonstrate that people unaware of the notification sounds (e.g., most customers in retail stores) are clearly outperformed on a wide range of behavioral metrics by those who know that such sounds are present (e.g., store personnel). As such, this study offers initial empirical evidence for the thesis that notification sounds may help store employees detecting and potentially preventing ongoing thefts in the retail environment, but more applied studies are needed to establish the applicability, generalizability, and real-world impact of these findings.

5.2. Future Work

To verify these tentative conclusions, we need to get a more nuanced understanding of the suitability of various sound alerts for employees and regular shoppers; future experimental research should therefore test which specific sounds and alert types generate the most (least) favorable (unfavorable) affective responses. Further studies are also needed to find effects of background music on both the design of notification sounds and on their noticeability.

Moreover, considering the large body of research documenting detrimental employee outcomes as a function of noise exposure (e.g. [55, 56, 57]), future studies should examine the impact of continuous exposure to sound alerts on employees' job satisfaction, subjective well-being, and the quality of service given to customers to ensure that the implementation of a store soundscape intended to reduce shoplifting does not carry costly consequences on these crucial outcomes.

6. CONCLUSION

In [9], we outlined a first proposal for design guidelines for short alert sounds in a retail environment. From the current work we conclude that by following these guidelines it was mostly the participants who were aware (knowing) of peripheral auditory notifications that also reacted to them. This may suggest that notification sounds will have different effect on those aware of a surveillance system (employees, shoplifters, and attentive purchasing customers), and those unaware (most regular customers).

An interpretation of the results we found is that persons unaware of audio alerts indeed did not get distracted or felt annoyed by the sounds. This is clear from the low notification ratio, and not least from the interviews where none mentioned any kind of annoyance connected to sound in the experience.

We suggest that the findings here give a promising starting point for future studies and investigations focused on improving the auditory environments in physical stores. The ambition is to increase customer shopping experiences and working conditions for employees, and the main aim is to develop novel methods to prevent people from shoplifting. This remains to be validated through field experiments.

7. ACKNOWLEDGMENT

The research was funded by the Hakon Swenson Foundation and The Swedish Retail and Wholesale Council (Handelsrådet). The authors wish to thank Gustav Frisk Arfvidsson for his work with the VR environment.

8. REFERENCES

[1] A. Lindblom and S. Kajalo, "The use and effectiveness of formal and informal surveillance in reducing shoplifting: a survey in sweden, norway and finland," *The International Review of Retail, Distribution and Consumer Research*, vol. 21, no. 2, pp. 111–128, May 2011.

[2] Y. Yamato, Y. Fukumoto, and H. Kumazaki, "Proposal of shoplifting prevention service using image analysis and ERP check," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 12, no. S1, June 2017.

[3] F. J. McShane and B. A. Noonan, "Classification of shoplifters by cluster analysis," *International Journal of Offender Therapy and Comparative Criminology*, vol. 37, no. 1, pp. 29–40, 1993-03.

[4] R. P. Carlyon, "How the brain separates sounds," *Trends in Cognitive Sciences*, vol. 8, no. 10, pp. 465–471, Oct. 2004.

[5] L. M. Axon, B. Alahmadi, J. R. C. Nurse, M. Goldsmith, and S. Creese, "Sonification in security operations centres: what do security practitioners think?" *Workshop on Usable Security (USEC) at the Network and Distributed System Security (NDSS) Symposium 2018*, July 2018.

[6] T. L. Chartrand, J. Huber, B. Shiv, and R. J. Tanner, "Nonconscious goals and consumer choice," *Journal of Consumer Research*, vol. 35, no. 2, pp. 189–201, 2008.

[7] T. Otterbring and P. Shams, "Mirror, mirror, on the menu: Visual reminders of overweight stimulate healthier meal choices," *Journal of Retailing and Consumer Services*, vol. 47, pp. 177–183, 2019.

[8] T. Hermann, A. Hunt, and J. G. Neuhoff, *The sonification handbook*. Logos Verlag Berlin, 2011.

[9] G. F. Arfvidsson, M. L. Eriksson, H. Lidbo, and K. Falkenberg, "Design considerations for short alerts and notification sounds in a retail environment," in *Proceedings of the Sound and Music Computing Conference*, 2021.

[10] R. Jung and A. Butz, "Effectiveness of user notification in ambient soundscapes," in *Proceedings of the workshop on Auditory Displays for Mobile Context-Aware Systems at Pervasive*, 2005, pp. 47–56.

[11] R. Jung and T. Schwartz, "Peripheral notification with customized embedded audio cues," in *International Conference on Auditory Display (ICAD2007)*. Georgia Institute of Technology, 2007.

[12] R. Jung, "Non-intrusive audio notification in emotion classified background music," in *Proceedings of Meetings on Acoustics 159ASA*, vol. 9, no. 1. Acoustical Society of America, 2010, p. 050001.

[13] W. Wahlster, M. Feld, P. Gebhard, D. Heckmann, R. Jung, M. Kruppa, M. Schmitz, L. Spassova, and R. Wasinger, "The shopping experience of tomorrow: Human-centered and resource-adaptive," in *Resource-Adaptive Cognitive Processes*. Springer, 2010, pp. 205–237.

[14] R. Jung, L. Spassova, and G. Kahl, "Product-awareness through smart audio navigation in a retail environment," in *2011 Seventh International Conference on Intelligent Environments*. IEEE, 2011, pp. 186–191.

[15] E. R. Spangenberg, B. Grohmann, and D. E. Sprott, "It's beginning to smell (and sound) a lot like Christmas: the interactive effects of ambient scent and music in a retail setting," *Journal of Business Research*, vol. 58, no. 11, pp. 1583–1589, 2005.

[16] M. Bar, "The proactive brain: using analogies and associations to generate predictions," *Trends in Cognitive Sciences*, vol. 11, no. 7, pp. 280–289, July 2007.

[17] P. Larsson, D. Västfjäll, P. Olsson, and M. Kleiner, "When what you hear is what you see: Presence and auditory-visual integration in virtual environments," in *Proceedings of the 10th annual international workshop on presence*. Citeseer, 2007, pp. 11–18.

[18] B. Gygi and V. Shafiro, "The incongruency advantage for environmental sounds presented in natural auditory scenes." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 2, pp. 551–565, 2011.

[19] E. R. Hafter, A. Sarampalis, and P. Loui, "Auditory attention and filters," in *Auditory Perception of Sound Sources*, W. A. Yost, A. N. Popper, and R. R. Fay, Eds. Springer US, 2008, pp. 115–142.

[20] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the acoustical society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[21] K. G. Grunert, "Automatic and strategic processes in advertising effects," *Journal of Marketing*, vol. 60, no. 4, pp. 88–101, 1996.

[22] A. M. Treisman, "Verbal cues, language, and meaning in selective attention," *The American journal of psychology*, vol. 77, no. 2, pp. 206–219, 1964.

[23] A. Dijksterhuis and H. Aarts, "Goals, attention, and (un)consciousness," *Annual Review of Psychology*, vol. 61, no. 1, pp. 467–490, 2010.

[24] M. W. Eysenck and M. T. Keane, *Cognitive psychology: A student's handbook*. Taylor & Francis, 2005.

[25] J. Salmi, T. Rinne, S. Koistinen, O. Salonen, and K. Alho, "Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention," *Brain Research*, vol. 1286, pp. 155–164, 2009.

[26] E. M. Kaya and M. Elhilali, "Modelling auditory attention," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1714, p. 20160101, 2017.

[27] I. McGregor, A. Crerar, D. Benyon, and C. Macaulay, "Soundfields and soundscapes: Reifying auditory communities," in *International Conference on Auditory Display (ICAD2002)*. Georgia Institute of Technology, 2002.

[28] C. Wisneski, H. Ishii, A. Dahley, M. Gorbet, S. Brave, B. Ullmer, and P. Yarin, "Ambient displays: Turning architectural space into an interface between people and digital information," in *International Workshop on Cooperative Buildings*. Springer, 1998, pp. 22–32.

[29] P. Alexanderson, "Peripheral awareness and smooth notification: the use of natural sounds in process control work," in *Proceedings of the third Nordic conference on Human-computer interaction*, 2004, pp. 281–284.

[30] L. Rayleigh, "XII. on our perception of sound direction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, 1907.

[31] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, 2002.

[32] Z. H. Zhou, "Sound localization and virtual auditory space," *Project report of University of Toronto*, vol. 19, 2002.

[33] W. Thurlow and P. Runge, "Effect of induced head movements on localization of direction of sounds." *The Journal of the Acoustical Society of America*, vol. 42 2, pp. 480–8, 1967.

[34] Y. Iwaya, Y. Suzuki, and D. Kimura, "Effects of head movement on front-back error in sound localization," *Acoustical science and technology*, vol. 24, no. 5, pp. 322–324, 2003.

[35] C. Kim, R. Mason, and T. Brookes, "Head movements made by listeners in experimental and real-life listening activities," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 425–438, 2013.

[36] A. Kulkarni, S. Isabelle, and H. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2821–2840, 1999.

[37] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[38] W. O. Brimijoin, A. W. Boyd, and M. A. Akeroyd, "The contribution of head movement to the externalization and internalization of sounds," *PloS one*, vol. 8, no. 12, p. e83068, 2013.

[39] C. Valzolgher, G. Verdelet, R. Salemme, L. Lombardi, V. Gaveau, A. Farné, and F. Pavani, "Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues," *Neuropsychologia*, vol. 149, p. 107665, 2020.

[40] H. Bahu, T. Carpentier, M. Noisternig, and O. Warusfel, "Comparison of different egocentric pointing methods for 3d sound localization experiments," *Acta acustica united with Acustica*, vol. 102, no. 1, pp. 107–118, 2016.

[41] J.-M. Pernaux, M. Emerit, and R. Nicol, "Perceptual evaluation of binaural sound synthesis: the problem of reporting localization judgments," in *Audio Engineering Society Convention 114*. Audio Engineering Society, 2003.

[42] R. H. Gilkey, M. D. Good, M. A. Ericson, J. Brinkman, and J. M. Stewart, "A pointing technique for rapidly collecting localization responses in auditory research," *Behavior Research Methods, Instruments, & Computers*, vol. 27, no. 1, pp. 1–11, 1995.

[43] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert, "An interactive virtual-environment generator for psychoacoustic research ii: Collection of head-related impulse responses and evaluation of auditory localization," *Acta Acustica united with Acustica*, vol. 86, no. 6, pp. 1046–1053, 2000.

[44] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, perception, & psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.

[45] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *The journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 2188–2200, 1990.

[46] A. d'Astous, "Irritating aspects of the shopping environment," *Journal of Business Research*, vol. 49, no. 2, pp. 149–156, 2000.

[47] P. Vickers, "Ars informatica – ars electronica: Improving sonification aesthetics," in *Understanding and Designing for Aesthetic Experience: workshop at HCI 2005: The 19th British HCI Group Annual Conference*, 2005.

[48] W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological Psychology*, vol. 5, no. 1, pp. 1–29, mar 1993.

[49] P. Bergman, A. Sköld, D. Västfjäll, and N. Fransson, "Perceptual and emotional categorization of sound," *The Journal of the Acoustical Society of America*, vol. 126, no. 6, pp. 3156–3167, 2009.

[50] F. Avanzini, M. Rath, D. Rocchesso, and L. Ottaviani, "Low-level models: resonators, interactions, surface textures," *The Sounding Object*, pp. 137–172, 2003.

[51] X. Pan and A. F. de C. Hamilton, "Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape," *British Journal of Psychology*, vol. 109, no. 3, pp. 395–417, Mar. 2018.

[52] J.-C. Servotte, M. Goosse, S. H. Campbell, N. Dardenne, B. Pilote, I. L. Simoneau, M. Guillaume, I. Bragard, and A. Ghuysen, "Virtual reality experience: Immersion, sense of presence, and cybersickness," *Clinical Simulation in Nursing*, vol. 38, pp. 35–43, Jan. 2020.

[53] M. L. Eriksson, L. Pareto, R. Atienza, and K. F. Hansen, "My sound space: An attentional shield for immersive redirection," in *Proceedings of the Audio Mostly Conference*. ACM Press, 2018.

[54] M. Alcañiz, E. Bigné, and J. Guixeres, "Virtual Reality in marketing: A framework, review, and research agenda," *Frontiers in Psychology*, vol. 10, 2019-07.

[55] E. Sundstrom, J. P. Town, R. W. Rice, D. P. Osborn, and M. Brill, "Office noise, satisfaction, and performance," *Environment and Behavior*, vol. 26, no. 2, pp. 195–222, 1994.

[56] H. Jahncke, S. Hygge, N. Halin, A. M. Green, and K. Dimberg, "Open-plan office noise: Cognitive performance and restoration," *Journal of Environmental Psychology*, vol. 31, no. 4, pp. 373–382, 2011.

[57] T. Otterbring, C. B. Danielsson, and J. Pareigis, "Office types and workers' cognitive vs affective evaluations from a noise perspective," *Journal of Managerial Psychology*, vol. 36, no. 4, pp. 415–431, 2020.