A DUAL-TASK EXPERIMENTAL METHODOLOGY FOR EXPLORATION OF SALIENCY OF AUDITORY NOTIFICATIONS IN A RETAIL SOUNDSCAPE

Emma Frid[1,2], *Sandra Pauletto*[1]

[1]KTH Royal Institute of Technology
[2]STMS IRCAM
Stockholm, Sweden
emmafrid@kth.se
pauletto@kth.se

Baptiste Bouvier[3], *Matthieu Fraticelli*[4]

[3]STMS IRCAM – CNRS – SU
[4]Département d'études cognitives – ENS
Paris, France
baptiste.bouvier@ircam.fr
matthieu.fraticelli@ens.psl.eu

ABSTRACT

This paper presents an experimental design of a dual-task experiment aimed at exploring the salience of auditory notifications. The first task is a Sustained Attention to Response Task (SART) and the second task involves listening to a complex store soundscape that includes ambient sounds, background music and auditory notifications. In this task, subjects are asked to press a button when an auditory notification is detected. The proposed method is based on a triangulation approach in which quantitative variables are combined with perceptual ratings and free-text question replies to obtain a holistic picture of how the sound environment is perceived. Results from this study can be used to inform the design of systems presenting music and peripheral auditory notifications in a retail environment.

1. INTRODUCTION

This study is part of the research project *Sweet-sounding Shop Goods*, which aims to present new knowledge on how sound can be used in stores to facilitate the staff's continuous monitoring of the premises, as well as examine customers' acceptance of alternative monitoring methods. The current work builds on previous research on peripheral auditory notifications (smooth auditory notifications) in a retail environment, see [1, 2].

A central concept for this work is the notion of *auditory salience*. Salience (or saliency) describes qualities of a stimulus that attract our attention and make certain elements in a scene stand out relative to others [3]. Scholars have emphasised that the research area dedicated to auditory salience remains in its infancy, with a lack of agreed-upon behavioural paradigms and standardised frameworks [4, 5]. Little work has focused on salience in rich natural auditory scenes (some examples include [6, 7]). A study on auditory salience in a set of dynamic natural scenes was presented in [5], with findings indicating that the space of auditory salience is multidimensional (spanning loudness, pitch, spectral shape, and other acoustic attributes), nonlinear, and highly context-dependent.

One method of determining the saliency of a stimulus is to investigate its ability to grab a listener's attention when their focus is directed elsewhere, using a dual-task experiment. In such an experiment, two tasks - one primary and one secondary - are performed simultaneously, and the participants' performance for the second task is used to determine the saliency of a particular auditory stimulus [8]. Below, we describe the dual-task experimental design that we have developed to examine the saliency of auditory notifications in a retail soundscape.

2. PROBLEM DEFINITION

The study described in this paper aims to investigate how sound notifications should be designed to be perceived by target users (e.g., store clerks) in a busy environment, such as a store, while at the same time blending into the soundscape without disturbing customers. Notifications should be noticeable by those who are trained to recognize them (clerks), but ideally be unnoticeable by others (customers). We believe that the notifications should be musical in their nature, in order to blend in with existing music in the store. We hypothesise that the saliency of the notifications might be affected by pitch contour, sound level (relative to the ambient and music sounds), and timing (i.e. when notifications occur in relation to existing music). More specifically, we hypothesise that the tempo of the music, and whether the notifications occur on beat or off-beat, might affect perception.

3. PROPOSED METHODOLOGY

3.1. Experimental Task

The perceptual experiment is based on a "Sustained Attention to Response Task (SART)" task [9], as presented in the PsyToolkit [10, 11][1]. The SART experiment involves the withholding of key presses to rare targets (more specifically, one target in nine). In the original design described in [9], single digits (1-9) were presented visually over a time period of 4.3 min. Each digit was presented for 250 ms, followed by a 900 ms mask. Subjects were instructed to respond with a keypress for every digit, except for the occasions when digit 3 appeared, in which they should withhold a response. Digits were presented using one out of five randomly allocated font sizes to enhance the demands for processing the numerical values. The mask following each digit was in the form of a ring with a diagonal cross in the middle. Both the digit and the mask were presented centrally in white against a black background. The subjects were instructed to give equal importance to accuracy and speed and they could use their preferred hand when performing the task. Robertson et al. [9] found that the performance on the SART correlates significantly with performance tests of sustained attention, but not other types of attention, thereby supporting the view

---

[1] See also `https://tinyurl.com/muah7s58` and `https://tinyurl.com/34f45us6`

that the task indeed can be used as a measure of sustained attention [9]. The SART has been used in a range of different settings, e.g. to explore associations between attention and media multitasking [12], the effect of different types of sounds on cognitive performance, workload assessment and emotional responses [13], and the effects of soundscapes on children's cognitive performance [14].

3.2. Experimental procedure

The experimental procedure is divided into separate parts, presented on subsequent pages in a web-interface that uses the JavaScript framework JsPsych[2]. The different steps are described in detail below. *1) Research subject information*: The participant will first be presented with an information page that introduces the project and the purpose and nature of the data collection. *2) Consent form*: The participant is then asked to reply to an online consent form in which they can agree or disagree to their data being collected as described on the previous page. *3) Demographic information*: This page presents questions concerning demographic information (age, gender identity, musical expertise, handedness[3], and brand of the headphones that are used). *4) Hearing test*: This step ensures not only that the participant has normal hearing; it also allows us to make sure that the headphones used are of adequate quality, i.e. that they can reproduce the frequency ranges required. The participant will be presented with a test sound and asked to set the level to a comfortable level. Since the volume is not intended to be changed after this step, these test sounds will be adjusted in relation to the stimuli sound level used in the actual experiment, described below. The hearing test then involves listening to a set of tones at different frequencies, counting them, and answering how many tones that could be heard in a number box. If the participant fails the hearing test, they will not be allowed to continue to the next step. The hearing test will only serve as a sorting mechanism to decide which participants that may proceed to the next steps (performance data from the hearing test will not be saved). *5) Practice trial:* Once the initial steps above are completed, the participant will proceed to a training session. In this step, the participant will perform the same SART task as in the main experiment, but different stimuli will be used. The task will be performed while listening to music mixed with a store soundscape sound, in which a set of auditory notifications will be presented at randomized time points. The participant will be prompted with the following instructions:

> *"You will have to perform a task requiring you to remain focused for a few minutes. Digits will be presented to you successively (1 to 9). Each time when a digit shows up, it will quickly be masked by a circle with a cross. Your task is to press the SPACEBAR in response to each digit, except for when the digit is a 3. Each digit is followed by a circle with a cross, which you can ignore. Meanwhile, a recording of a store ambiance where there is music playing will be diffused."*

To explore the effects of a fictitious customer versus staff perspective, participants will then be randomly allocated to two groups (customer (C) versus staff (S) scenario), with different continuations of the instructions:

> C: *"In this soundscape, various sounds can occur. As you will be focused on the task you have to perform, it could happen that a sound disturbs you. In that case, just press the X button so that we know it happened and continue the task."*

> S: *"In this soundscape, notification sounds will be played when customers are interacting with products or something unusual is happening in the store. In this context, a notification sound is a sound that is short in length, that serves the purpose of notifying the staff in the store about events taking place. When you think that you heard a notification, just press the X button so that we know you got the information and continue the task."*

Participants will be instructed to give equal importance to the speed and accuracy of their responses. *6) Test trial - music 1:* The participant will be informed that there will be two test sessions, each with different background music. The task is the same as for the practice trial, but the duration will now be slightly longer (6 min). For both test trials, the system will randomly select between two music stimuli: fast music (120 BPM) and slow music (80 BPM). Two different types of music at different tempi were chosen since we hypothesised that the effect of perceiving a sound on versus off-beat might depend on tempo. *7) Break:* After completing the first test trial, the participant will be encouraged to take a short break. *8) Test trial - music 2:* The second test trial follows the procedure of the first test trial, the only difference is the music: if 120 BPM was used in the previous session, the music with 80 BPM will be used, and vice versa. The following data will be collected during the test trials: SART digits with corresponding timestamps, and key presses with respective time stamps. *9) Questionnaire*: The experiment will conclude with a questionnaire focused on the auditory notifications used in the study. To remind the participant about the different types of sounds used, notifications in each timbral category will be replayed. The participant is asked to listen to the notifications (as presented with the fast versus slow music) and rate how much they agree with the following statements, using 10-point Likert scales ([strongly disagree: 1, strongly agree: 10]):

1. *The notifications are pleasant.*
2. *The notifications communicate a sense of urgency.*
3. *The notifications fit well with the context, i.e. with the store ambience and the music played in the background.*
4. *Listening to these notifications many times (repeatedly) makes me irritated/annoyed.*

This is followed by two multi-line free text questions:

5. *Do you think that these notification sounds would fit in a store context? Please motivate why you think they would fit, or why you think that they would not fit.*
6. *Do you have any other comments about these notification sounds, or how they blended in with the background music?*

The above-outlined steps are repeated for three timbral notification categories (*instrumental, electronic,* and *reverberant*, see Section 3.3) and the slow, as well as the fast, music. *10) Concluding questions:* The final page of the experiment includes two questions in free-text format:

1. *Was the task performed in this study difficult to perform? If so, please describe what was difficult.*
2. *Is there anything else that you would like to share with us?*

3.3. Stimuli

A sound designer was commissioned to compose the music and the notifications. They were asked to create "typical" store music that should be pleasant, but not particularly original to attract attention, and with no unexpected dynamics that, again, could attract attention. The music is available as supplementary material[4]. The sound designer was asked to create notifications with a musical nature with an emphasis on a recognisable instrument, electronic

---

[2] `https://www.jspsych.org/7.3/`, see also `https://matthieufra.github.io/jsPsychPDS/`

[3] The participant will be asked to use both hands when performing the tasks.

[4] See `https://tinyurl.com/45ed46ex` (password: icad)

sound, and a reverb effect (referred to as *instrumental*, *electronic*, and *reverberant* above). The notifications should not be longer than 2 s and for each category, there should be a version with ascending pitch and a version with descending pitch. The music and the 6 notifications were created in Logic Pro. We selected a store ambience after consulting a series of sound effect libraries and comparing what we found with ambiences in stores that do not usually play music. We wanted an ambience with no recognisable speech, that remained dynamically similar throughout the duration of the music. We selected a store ambience from the BBC Sound Effects Library [5] which was dynamically appropriate but contained some speech, and subsequently edited the speech out of the ambience, using Adobe Audition.

Randomised on beat and off-beat time markers for the notifications were generated for both fast and slow music, using Supercollider. This allowed for random distribution of notifications along the duration of the music, ensuring that the notifications were presented with an average time difference of 10 s (n=36), for a total of 6-min music, regardless of tempo. The timbral notification categories were presented in one configuration with an upward pitch, and one with a downward pitch, giving a total of 6 unique notification sounds. All notifications were also presented 3 times on versus offbeat, giving 36 notifications in total. The SuperCollider script made sure that notifications that should be "offbeat" were not presented on half beats, or too close to an actual beat, since this could result in the notification being perceived as "on beat". A total of 10 different randomizations were generated and saved as multitrack audio files, along with metadata files (CSV files) with logging of time stamps and notification characteristics.

The final stimuli were created by mixing music, ambience and notifications in Adobe Audition. Firstly, all separate sounds were normalised to -23 LUFS in Audacity. In the mix, the music was left at 0dBs, while the ambience was set to -6dBs, and the notifications set to -11dBs. The volume of the ambience track was lowered in relation to the music level to match relative levels found in similar store soundscapes found in different sound effects libraries. The notifications were lowered on the basis of a pilot study conducted with 9 participants at [removed]. The findings from the pilot (see Section 3.6) suggested that -11dBs was the average threshold at which participants would not be able to detect notifications anymore.

3.4. Participants

Participants will be recruited from university staff and students at KTH Royal Institute of Technology. A minimum of 60 participants will be recruited (30 for staff versus customer scenario/instruction). Apart from having access to a laptop with headphones, the only prerequisite for participation will be to have normal hearing (with or without the aid of a hearing device). This will be informally tested using the hearing test described in Section 3.1.

3.5. Analysis

For each participant and test session (randomly assigned to a "staff" versus "customer" scenario) collected data will be processed to compute reaction times and accuracy metrics. Statistical analysis will be performed to evaluate the effect of sonic properties of the auditory notifications (pitch going up/down, timbral properties: instrumental, electronic, reverberant) on performance metrics and perceptual ratings, for fast versus slow music. Qualitative data in free text format will be analysed using content analysis.

3.6. Pilot study

We conducted a pilot study to identify at what volume the notifications should be set, in relation to music and ambience, so that most of them could be perceived in the absence of a distracting task. 9 participants were recruited (7 M, 2 F, average age 31.8 yrs, SD 8.7). Music was set to 0dBs, and ambience level to -6dBs. The sounds were played from a laptop in a quiet studio environment. The participants used headphones. 4 participants first listened to the slow music, and 5 participants listened to the fast music first. The notifications were initially set to -30dBs so that they could not be heard. While the subjects were listening to the music and the ambience, the experimenter was able to increase the volume of the notifications by 3dBs. The experimenter would wait for 4 to 6 notifications to play before increasing the volume again. The participants were instructed to put their hand up or say "yes" when they started hearing something that they did not recognise as either music or ambience. They had their eyes closed throughout the experiment, so that they could not see the experimenter controlling the volume fader. The experimenter annotated the level at which participants started hearing something (indicated by "FIRST") as well as when they started hearing all the notifications (indicated by "ALL"). Collected results were averaged across participants. Findings indicated that the participants started hearing all notifications when these were set at about -11dBs (± 1-2 dBs) compared to the music. A pilot of the final test will confirm this with more precision.

4. CHALLENGES AND PROPOSED SOLUTIONS

The *choice of notifications* is an important design decision that may influence if the sounds are perceived as alarms or subtle notifications. The notifications should communicate an appropriate level of urgency. A complicating factor in this context is that the auditory notifications are superimposed on the music that is played. It is important that the sounds are not only informative but also fit in with the sonic branding and the music that is played. In the current work, we tried to limit the effects of confounding variables by making sure that all notifications had the same duration and similar pitch contours. However, restricting a sound designer in such a way in a real-world setting might limit their musical expression too much. Another complicating factor in this context is *masking* - different notification sounds will be more or less masked by the music and the ambient sounds of the store setting. Masking levels will also vary throughout the musical piece. One way to address this could be to use computational tools to analyse masking levels based on frequency or loudness at various time points and to adjust notification properties (e.g. bandwidth) accordingly. In our current work, we addressed the masking issue by listening to many different notification sounds, and selecting sounds that we perceived stood out against the sonic backdrop. For example, we did not use sounds with strong noise components, since they were easily masked.

Another challenge is the temporal *randomization* of auditory notifications in relation to the metric structure of the music. A number of questions relating to what is actually *perceived as "on beat"* emerged when coding the randomization algorithm. For example: how many milliseconds are required for a notification to be robustly perceived as "offbeat"? A too small time difference, in which the notification is presented just after the beat, might accidentally result in the sound to be perceived as "on beat", due to the

---

[5] `https://sound-effects.bbcrewind.co.uk/search?q=07026105`

temporal resolution of hearing. Monoaural gap detection thresholds measured for hearing are usually in the millisecond range, but such experiments do of course not take aspects of musical structure into account. Moreover, it can be discussed whether temporal instances close to a half-beat should really be considered as "offbeat"; half beats are not usually interpreted as being perceptually off, musically. Yet another complicating factor in this context is dominant beats and accents in different types of music. In our work, we attempted to address most of the above-described issues by putting constraints on the randomization algorithm so that all time differences were clearly perceived as offbeat when they were supposed to be offbeat. This was done by adding or subtracting appropriate milliseconds, and fine-tuning the results through repeated listening.

Finally, it should be noted that asking participants to press "X" if they heard a notification may result in *divided attention*; some attention might be directed to the fact that this kind of sound might occur, and some top-down processing might thus be involved. Ideally, in order to only evaluate bottom-up processing, i.e. the attention captured by the notifications, one should not mention that there will be notifications. However, if the main task is sufficiently demanding, participants should quickly stop paying attention to what they hear, meaning that they can be fully engaged in the SART task, reducing the risk of divided attention.

5. DISCUSSION

Given that the task of exploring the influence of different sonic attributes on attention of notifications in such a complex setting as the one presented in a store soundscape is not a trivial task, one may ask if the findings from an online experiment such as the one described above actually can be applied to a physical store scenario. Although the above-described experiment is neither a controlled lab study nor an ecologically valid experiment performed in an actual store, we still believe that the findings can be useful. In particular, the main advantage of the proposed method is that it allows us to obtain preliminary insights about the effectiveness of specific sound designs prior to conducting experiments *in the wild*. This may save both time and resources, and findings may guide future design processes. If - with an adequate sample size - the results obtained from the described experiment indicate no significant difference between notifications presented on versus offbeat, this would suggest that notifications can be programmed to be played at any time, directly after an event in the store. As such, there will be no need to quantize the sounds to happen on the beats of the music. However, if there is a significant difference, this implies that the timing of the notifications might be of certain importance and that the system providing notifications should take synchronization of auditory events into account.

6. CONCLUSIONS

This paper proposes an experimental design of a dual-task experiment aimed at exploring salience of auditory notifications when presented against a retail soundscape. The main task used in this experiment is a Sustained Attention to Response Task (SART) to be performed in a web browser, while listening to a store soundscape. The proposed methodology can be used to examine the effect of sonic properties on auditory saliency, which in turn may guide the design of auditory notifications used for peripheral notification in stores.

7. ACKNOWLEDGMENT

The research was funded by the Hakon Swenson Foundation and The Swedish Retail and Wholesale Council. The authors would like to thank Håkan Lidbo, who designed the notifications and composed the music.

8. REFERENCES

[1] K. Falkenberg, E. Frid, M. L. Eriksson, T. Otterbring, and S.-O. Daunfeldt, "Auditory notification of customer actions in a virtual retail environment: Sound design, awareness and attention," in *International Conference on Auditory Display (ICAD)*, 2021.

[2] G. F. Arfvidsson, M. Ljungdahl Eriksson, H. Lidbo, and K. Falkenberg, "Design considerations for short alerts and notification sounds in a retail environment," in *Sound and Music Computing Conference (SMC)*, vol. 2021, 2021, pp. 261–267.

[3] S. Treue, "Visual attention: the where, what, how and why of saliency," *Current opinion in neurobiology*, vol. 13, no. 4, pp. 428–432, 2003.

[4] 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.

[5] N. Huang and M. Elhilali, "Auditory salience using natural soundscapes," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 2163–2176, 2017.

[6] K. Kim, K.-H. Lin, D. B. Walther, M. A. Hasegawa-Johnson, and T. S. Huang, "Automatic detection of auditory salience with optimized linear filters derived from human annotation," *Pattern Recognition Letters*, vol. 38, pp. 78–85, 2014.

[7] O. Kalinli and S. S. Narayanan, "A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech." in *INTERSPEECH*, 2007, pp. 1941–1944.

[8] V. Duangudom and D. V. Anderson, "Identifying salient sounds using dual-task experiments," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2013, pp. 1–4.

[9] I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend, "'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects," *Neuropsychologia*, vol. 35, no. 6, pp. 747–758, 1997.

[10] Stoet, G., "PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments," *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, 2017.

[11] G. Stoet, "PsyToolkit: A software package for programming psychological experiments using Linux," *Behavior Research Methods*, vol. 42, pp. 1096–1104, 2010.

[12] B. C. Ralph, D. R. Thomson, P. Seli, J. S. Carriere, and D. Smilek, "Media multitasking and behavioral measures of sustained attention," *Attention, Perception, & Psychophysics*, vol. 77, pp. 390–401, 2015.

[13] K. Kallinen and J. Gylden, "The effects of sound interference on soldiers cognitive performance, workload assessment and emotional responses," in *International Virtual Conference on Human Interaction and Emerging Technologies*. Springer, 2022, pp. 974–980.

[14] S. Shu and H. Ma, "Restorative effects of classroom soundscapes on children's cognitive performance," *International Journal of Environmental Research and Public Health*, vol. 16, no. 2, p. 293, 2019.