SONIFICATION OF TWITTER HASHTAGS USING EARCONS BASED ON THE SOUND OF VOWELS

Emil MYRESTEN (myresten@kth.se), **David LARSSON HOLMGREN** (ndtho@kth.se), and **Roberto BRESIN** (roberto@kth.se) (0000-0002-3086-0322)

KTH Royal Institute of Technology, Stockholm, Sweden

ABSTRACT

The amount of notifications we receive from our digital devices is higher today than ever, often causing distress in users constantly having to move their devices into the center of attention, digesting the information received visually. In this study we have tested the use of short sound messages, earcons, in order to notify users about the arrival of different Twitter messages. The idea was to keep the arrival of new messages in the periphery of attention while simultaneously monitoring other activities. Using Twitter hashtags as the underlying data, a sonic abstraction was made by mapping the vowels present in a hashtag to a melody, and by enhancing the formant frequencies of these vowels. This gives rise to the question of whether enhancing vowel presence through formant synthesis aids the implicit learning of earcons, with Twitter hashtags as the underlying text data. A methodology is described, with a mapping of several phonetic vowels containing the fundamental frequency, f0, the first two formant frequencies, f1 and f2, for each vowel, as well as a rhythmic mapping based on the hashtags' syllables as well as where the emphasis lies. An application was developed to receive tweets in real time using real data from Twitter, playing the earcons associated with hashtags of actual Twitter messages. Results of a user test show that participants were able to recognize the earcons related to each of the hashtags used in the experiment to a certain degree.

1. INTRODUCTION

The use of audio for notifying users of new incoming information has seen an increased use in recent years, as we have moved towards a more digital society. One of the several possible ways to communicate by means of audio is through the use of earcons, a term derived from its visual counterpart, icons, or 'eye-cons' [1]. Earcons are generally defined as non-verbal musical audio messages, meant to provide information to the user, without any particular connection to the information carried [2]. In other words, they are abstract representations of data, that you need to be exposed to before you can understand them properly. This gives earcons as audio notifications the ability to deliver messages in public, without necessarily sharing the information to people that are unfamiliar with the meaning of the earcon.

Notifications from smartphone apps represent a common type of everyday earcons, but despite their latent ability to convey the information of the notification, they often tend to sound similar and generic, making it difficult to differentiate between the content associated to each earcon, and requiring users to assimilate the information visually. This type of genericness can be addictive, especially when the content of notifications differs from case to case while the earcon remains the same [3], and it forces the smartphone into the center of our attention. This approach doesn't take advantage of the possibility for earcons to make use of the cocktail party effect [4]; our auditory senses allow us to be selective of what we pay attention to, even when exposed to multiple streams of stimuli. Earcons have the potential to be implemented as calm technology [4], enabling users to receive information in a subtle and unobtrusive way without interrupting other ongoing activities.

Earcons are part of our everyday soundscape, and a large part of them come from our smartphones. A reason why the current general use of earcons could appear distant from being a calm technology might be that the revenue models used by many of the big tech corporations are based on users' activity on the platform in question [5]. The more time the user spends on the platform, the more revenue there is to be gained, and the push notifications sent out by these platforms are often accompanied by an ambiguous earcon, a common way to steer user activity towards a platform by demanding the users' attention. This constant interruption of our attention makes us distracted and hyperactive, which in turn affect our productivity and mental health [6]. However, in Fitz's et al. [3] study of batching of notifications, it was also evident that not being notified created inattention, stress and anxiety, one of the reasons being that many notifications are useful and necessary, and missing out on them was causing distress. Thus, by adjusting the way we receive notifications, it's possible to enhance the psychological well-being of the user.

The way we produce different speech sounds relies on frequencies that resonate in the vocal tract, called formant frequencies [7]. The first three of these formant frequencies characterize each vowel sound that we produce, such as the 'a' in 'apple', and it has been shown that these three formants play a large enough part in speech perception, that a spoken word or sentence can remain comprehensible when recreated as a synthesized sinusoidal sound sig-

nal based on these three frequencies, as long as some pre-existing knowledge of the content exist [8]. Using formant frequencies to recreate vocal sounds in synthesized audio is called formant synthesis, and by applying formant synthesis to earcons, we aim to examine the effect of formant synthesis on implicit learning of earcons.

2. BACKGROUND

2.1 Hashtags and sound notification in Twitter

Twitter is a social media platform with millions of users worldwide that operates under the revenue model mentioned in 1 [9]. Users can send short Twitter posts, so-called tweets, that other users can respond to, share and retweet, through their own Twitter profile.

A large part of what makes Twitter into what it is, is the hashtag, a metadata tag consisting of a keyword prefixed by a pound sign, #. The hashtag acts as a keyword, and it is created and assigned by users to their tweets, which then links the post to a certain context. It works as a system to categorize the tweets, enhancing their searchability and visibility [10, 11]. Hashtags are often short words or sentences that act as a sort of umbrella term for messages that relate to each other. Since these short text snippets, in themselves, contain a lot of meaning, we consider hashtags to be suitable as the underlying data to earcons.

Twitter's current default notification sound is a short whistle, reminiscent of the twitter of a bird. It is used to increase user awareness of receiving a tweet [12], but it doesn't take full advantage of the information carrying properties of sound and music communication. Earcons could be used for delivering more complex and comprehensive messages in a very short time and could effectively allow users to identify tweets.

2.2 Earcons

McGookin et al. [2] mention several different categories of earcons, and this paper will focus on two of them; transformational and compound earcons. Transformational earcons are constructed with the use of a set of rules where certain parameters of the underlying data are mapped to certain acoustic and/or musical attributes, such as timbre, register, pitch, and rhythm. The authors describe a compound earcon as an earcon consisting of multiple earcons, much like a sentence consisting of multiple words. According to the authors, the recommended time gap between the earcons in a compound earcon is at least 0.1 s, to enable users to distinguish them as separate earcons and at the same time bound to each other. By using transformational and compound transformational earcons and an intuitive mapping [2, 13], we can build complex informative sound notifications. To strengthen the likelihood for implicit learning of the earcons to occur, the use of formant frequencies, presented in the next section, could present a useful mapping tool.

2.3 Formant frequencies

When we speak, sound produced by the vibration of our vocal cords has to travel through the vocal tract, the section from our vocal cords to our lips that acts as a resonator for multiple frequencies [14]. These frequencies are called formants, named `f0`, `f1`, `f2` and ascending. The frequencies `f1` and `f2` are what distinguish different vowel sounds from each other, while `f0` dictates the pitch of the sound. Thus each vowel that can be produced by the human voice is connected to a specific, but not unique, setting of the articulatory apparatus, a child can produce the same vowel sounds as its father, despite not being able to set the articulatory model to the exact same parameters; the fathers' vocal tract is likely to be a lot larger.

3. METHOD

3.1 Designing the earcon

For the purpose of this study, a set of transformational earcons was created (see Section 2.2). An overview of the mapping used for the earcons is presented in Table 1.

To perform the study, twelve hashtags were chosen, each of varying degrees of activity. These are listed in Table 2.

As suggested by Walker and Kramer [13], we chose to categorically map each phonetic vowel to a specific note based on the `f1` of the phonetic vowel, using a C-major

Data parameters	**Auditory attributes**
Each phonetic vowel is given their own fundamental frequency, f0.	Pitch: The notes of the C-major scale.
The two first formant frequencies, f1 and f2, in the phonetic vowel.	Frequency: Amplification of f1 and f2 in the musical note, f0, mapped from the phonetic vowel.
The syllables & emphasized vowels of the hashtag	Rhythm: The syllables are represented by individual notes, with emphasized vowels represented by eighth notes and the unemphasized vowels by sixteenth notes.

Table 1. A table representing the earcon set's mapping of data paramaters to auditory attributes.

Word	**Vowel sounds**
#spring	/i/
#march	/ɑ/
#facebook	/æ/, /ʉ/
#coffee	/ɔ/, /i/
#stockholm	/ɑ/, /ɔ/
#joebiden	diphthong: /o//ʊ/, diphthong: /a//i/, /ɛ/
#donaldtrump	/ɑ/, /a/, /ɵ/
#crowdfunding	diphthong: /a//ʊ/, /ʌ/, /i/
#influencer	/i/, /ʉ/, /æ/, /æ/
#influencermarketing	/i/, /ʉ/, /æ/, /æ/, /ɑ/, /ɵ/, /i/
#happybirthday	/ɑ/, /y/, /œ/, diphthong /e//y/
#fridaysforfuture	diphthong /a//i/, diphthong /e//y/, /o/, /ʉ/, /ɛ/

Table 2. Selected hashtags together with the phonetic vowels present in each word.

scale starting from C3 and ascending. The polarity of the mapping is deemed intuitive, since the vowel with the lowest `f1` frequency was assigned to the first tone of the musical scale, and the rest following the same rule in ascending order. This was done to prevent, as often as possible, `f0` from reaching frequencies above `f1`, in order to maintain as much of the vowel characteristics as possible in the resulting sound, while still having a clear difference between the different phonetic vowels. Although, this was not possible for every vowel, as evident by Table 3.

From the 12 hashtags chosen for the test, 15 different phonetic vowels were extracted, and thus the `f0` mapping stretched from C3 to C5. The frequency values used in the mapping of formant frequencies `f1` and `f2` of each phonetic vowel was based on multiple studies [14–16]. The Pure Data-patch *Madicken*, a formant synthesis generator by Anders Friberg, was used to further match the perceived vowel sounds of the Interactive IPA Chart [16] to their respective `f1` and `f2`. The final mapping of each phonetic vowel is visualised by Table 3.

The rhythm and length of each earcon was decided based on how many syllables the hashtag had. We chose to map the emphasised syllable of each hashtag as an eight note, whereas other syllables were represented as sixteenth notes. The emphasized syllable was subjectively decided based on our own pronunciations of the word, being Swedish English speakers, and when in doubt, the Merriam Webster online dictionary was used as a guideline. This creates a rhythm intuitively based on the pronunciation of the word. If there were multiple words in a hashtag, a sixteenth pause was inserted between the words. Using this method, we could also easily create compound earcons, with `#influencermarketing` being the relevant one in this study. All earcons were produced at 96 bpm, which would make this pause significant enough to be noticed [2]. The choice of using simple rhythms could help in focusing users' attention on the sound of the sonified vowels, rather than on complex rhythmical patterns.

We chose flute as the timbre for the melody, as its length and shape resembles the human vocal tract. Its frequency spectrum is also suitable for the frequency response of smartphone speakers [17]. A musical timbre is also the preferred choice in facilitating learning according to existing design guidelines for earcons created by Brewster et al. [1]. The hashtags were mapped into melodies using the methods described above, and `f1` and `f2` were amplified by 10 dB each, for each vowel sound, by applying an parametric EQ onto the melodies, using the EQ plug-in built into Logic Pro X. In the case of diphthong vowels, that is, vowels that glide from one phonetic vowel to another, a time varying filter was applied, and thus the starting `f1` and `f2` were different from the ending `f1` and `f2` of the syllable. An example would be `#joebiden`, where 'oe' constitutes a diphthong vowel. In these cases, `f0` was chosen to be that of the first vowel in the diphthong. In cases where `f1` was of lower frequency than `f0`, the `f1` frequency was moved to a frequency slightly above `f0`, in coherence with previous research [14].

Vowel	/y/	/u/	/ʉ/	/i/	/o/	/e/	/ø/	/ɑ/	/ɔ/	/œ/	/ʌ/	/ɛ/	/ɑ/	/a/	/æ/
f1	223	283	321	375	433	450	482	482	557	587	673	720	745	958	970
f2	1916	620	1544	2700	584	2250	1124	1292	764	1280	1148	2000	1064	1412	1760
Note	**C3**	**D3**	**E3**	**F3**	**G3**	**A3**	**B3**	**C4**	**D4**	**E4**	**F4**	**G4**	**A4**	**B4**	**C5**
Freq	262	294	330	349	392	440	494	524	588	660	698	784	880	988	1048

Table 3. Phonetic vowels and frequency mapping (in Hz).

3.2 The web application

A web application was developed to display tweets as they were received in real-time using Twitter's API, which was configured to only load tweets posted under the hashtags selected for the study (see Table 2). Each time a tweet was received, the corresponding hashtag was displayed on screen and its associated earcon was played. If multiple tweets from the same hashtag were received in succession, the earcon would only be played once. This decision was made to keep the number of earcons played by the web application to a reasonable amount. The average amount of tweets posted to the twelve hashtags was estimated to be roughly 3 tweets a minute. The web application was designed to gather data about how many times a user visited the website, as well as the amount of encounters each user had with any specific hashtag/earcon.

It has been shown that an earcon can be recognized with up to 90% accuracy after just 5 minutes of training [2], and we estimated that our participants would see less than 5 minutes of active training for each hashtag over the span of the 75 minutes supposed to be spent by users, as explained in the next Section 3.3. Furthermore, despite verbal training being proven to be the most effective when learning earcons, a real-life scenario such as this was used to study the implicit learning of the earcons.

No personal data was gathered by the web application.

3.3 Training, test and interview

19 participants (8 F, 11 M), with normal hearing, took part in the experiment. Each participant was provided with a personal username to be used with the web application and instructions on how to connect to it. They were instructed to visit the web application for 15 minutes daily, for five days in a row, where tweets would be received in real time. They were allowed to use the web application whenever they wanted to, and thus the amount of encounters with any specific earcon was assumed to be unique to each participant. Further, they were allowed to perform other activities during training, as long as they could hear the earcons, and were instructed to note the hashtag displayed by the web application whenever an earcon was played. This method was chosen in order for learning to take place while participants were attentive elsewhere, while still providing feedback to reinforce the sound-hashtag relationship. It was deemed suitable to use real-time data, since it provides insight into how implicit learning would take place during real use. With tweets being received in real-time, the number of times each earcon was heard was different for every participant. Participants were allowed to use whatever playback device they had at their disposal.

The test was constructed to check if the mapping of formant frequencies was recognized by the participants. To

do this, two reference earcon sets were produced as complements to the one designed using the method described in 3.1. One of these complementary earcon sets was made without any specific mapping, while the other had the same `f0` mapping as the one used during the week, but without `f1` and `f2` amplified. This was done to test how well participants could distinguish the earcons where the formant synthesis was applied, as a way to understand the difference between the simple conditioning of an earcon melody, that has been proven to work before [2], and the formants themselves as an independent means of sound characteristic. The three different earcon sets, organized according to hashtag, are available online for listening [1] .

After the training, participants were invited to an individual meeting in Zoom [2] , where the test was conducted. For each hashtag, the three earcons were played in random order. Participants were only allowed to listen to each earcon once, in order for the test to reflect a real-use scenario. They were then asked which of the three earcons they associated with the hashtag in question.

A short interview was conducted, where participants were asked what playback device they used during both training and the listening test, if they had any musical background and if they relied on audio notifications when using their digital devices.

4. RESULTS

4.1 Test results and training data

The main result of this study is that participants were more likely to choose earcons from the sets with mapping, both with and without formant synthesis, but that it can not be strengthened at a significance level of 5% that the formant synthesis makes any difference in how the participants perceive the earcon [3] .

On average participants selected the formant-based earcon 4.7 (SD=1.8) times out of 12 possible. When looking at both the formant-based earcon and the earcon with only a melody mapping, the average selection rate per participant was 9.2 (SD=2.5). The selection made correlates slightly to the amount of encounters a participant has had with the earcon in question, as can be seen in top of Figure 1. Furthermore, the amount of non-mapped earcons selected is not increasing with the amount of encounters, and upon removing one-syllable earcons such as `#spring` and `#march`, that participants reported to be especially difficult to distinguish between the mapped and non-mapped version, the amount of wrong answers decrease with increased amounts of encounters (even though this decreases the significance of the correlation).

In the study, the amount of encounters per hashtag varied, with `#coffee`, `#stockholm`, `#fridaysforfuture`, `#influencermarketing` being the lowest, ranging between 0 and 9. In spite of this, `#stockholm` was one of the hashtags with the largest

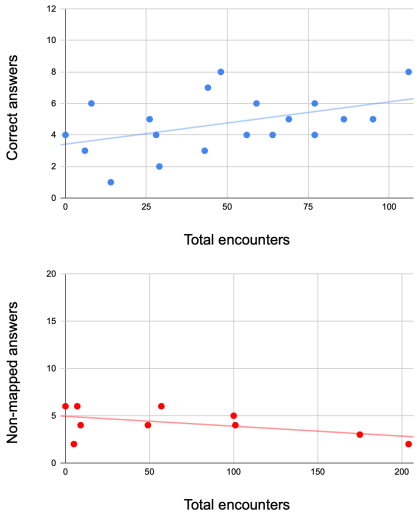

Figure 1. Top: Number of recognized earcons vs. number of encounters (Correlation coefficient R=0.45). Bottom: Number of recognized non-mapped earcons, not including single-note earcons `#spring` and `#march`, vs. number of encounters (Correlation coefficient R=-0.58).

amount of participants selecting the formant-based earcon, but the others saw an equal distribution of answers.

58% of the test participants answered that they have a musical background in some sense. The mean of correct [4] answers for participants with a musical background (n=11) was 4.5 (SD=2.0) while it was 4.9 (SD=1.5) for those without musical background (n=8). Data show a steeper decreasing trend for choosing the non-mapped melody in participants with a musical background, but also show that the participants with musical background and no encounters were more inclined to choose the non-mapped earcons. Furthermore, 65% of participants used speakers during the training, whereas 35% used headphones. During the test, 32% of the participants using speakers, while 68% used headphones. As visible in Figure 2, participants who use headphones during the test seem to be performing better at a lower amount of encounters, but that the discrepancy diminishes with a higher amount of encounters (correlation coefficient R=0,42). The participants who used the same playback device during both training and test sessions identified a increasing amount of formant-based earcons, with the headphone-users performing better.

4.2 Statistical analysis

The collected data proved to be approximately normally distributed, using a Shapiro-Wilk test [5] , for mapped earcons [6] whereas the answers for non-mapped earcons were not. These were therefore analyzed using a two-sided one-sample Wilcoxon Signed Rank test to investigate if it was less likely for participants to choose an earcon

[1] `https://doi.org/10.5281/zenodo.4761367`
[2] `https://zoom.us`
[3] Training data, detailed test results, plots, and link to earcons used in this study can be found in the appendix material: `https://zenodo.org/record/5645336`
[4] Selections of either the formant-based earcon or the earcon with only the melody mapping.
[5] All statistical tests described in section 4.2 were conducted in IBM SPSS Statistics (`https://www.ibm.com/products/spss-statistics`)
[6] With and without formants enhanced.

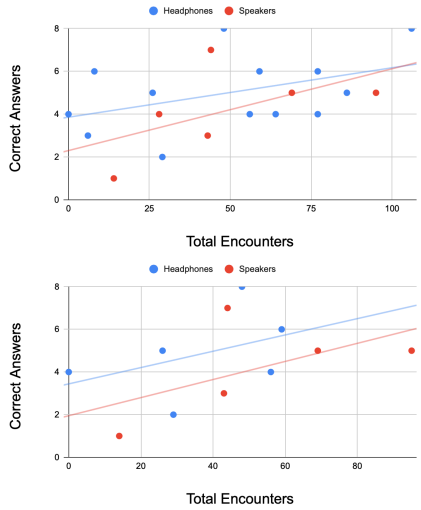

Figure 2. Top: Number of recognized earcons when using playback devices only during the test (Correlation coefficients: $R_{Headphones}$=0.42, $R_{Speakers}$=0.57). Bottom: Number of recognized earcons when using the same playback device during both training and test (Correlation coefficients: $R_{Headphones}$=0.43, $R_{Speakers}$=0.55).

with a previously unheard melody, i.e. a non-mapped earcon, than those from the other two earcons sets. The null hypothesis for the non-parametric test is that the observed median of the amount of answers with the non-mapped earcon chosen is equal to the hypothesized median 4. Calculations show that the test statistic is equal to 34.50 (SE=20.83) and the observed median is equal to 2 (p = 0.044), which means the null hypothesis can be rejected, and thus the conclusion that participants are more likely to choose the earcons from the mapped earcon sets rather than the ones from the non-mapped earcon set, is drawn. The next question to answer is if it is more likely for the participants to choose the formant-based earcon set rather than the other alternatives. Since they were shown to be approximately normally distributed, a t-test is performed. The null hypothesis is that the mean of the amount of formant-based answers from each participant is less than or the same as 4. The test is proven to be significant at the level α=0.10 but not at the level α=0.05, which means the null hypothesis cannot be rejected. Thus, the conclusion that formant-based earcons are more likely to be recognized and remembered cannot be drawn.

5. DISCUSSION

The study at hand tested a new concept, where we examine the effect of formant synthesis and vowel enhancement in identifying earcons, potentially allowing users to understand the content they are notified about, without demanding that they digest the information visually. This would allow notifications to be received unobtrusively, with users continuing the current activity without being distracted.

Results show that, in coherence with previous research, participants were able to systematically choose the correct earcon melody, but that the effect of the enhanced formants cannot be strengthened. On average, we observed a trend showing that participants chose formant-based earcons more often than earcons with only the `f0` mapping. This partly confirms results from previous research [8] in which it was found that when substituting sounds of each vowel in a word with that of three sine waves with frequencies corresponding to its first three formants it was still possible to recognize the word.

Using a parametric EQ as formant enhancer may not have been the optimal choice. Even though it was easy-to-use and enabled the possibility of enhancing the gain of specific frequencies, it was still hard to recreate exact representations of the formant frequencies of the vowels in question, which could've affected the participants ability of distinguishing the formant synthesis. Other methods for formant synthesis could be explored in future studies, e.g. more sophisticated audio analysis for formant tracking [7].

One advantage of using concepts such as hashtags as the underlying data for sonification is that they already exist as an umbrella term, and no further analysis of the content is needed in order to convey the information to users. This is a suitable foundation to build earcons upon, and using formant synthesis to further sonify such data is interesting, since even in speech the hashtag itself is an abstraction of the information it represents. Applying formant frequencies and enhancing the vowel presence in hashtags could therefore be a way of effectively conveying the data to users, while still requiring the earcon to be learned.

Looking at the data for `#spring`, the 10 participants who selected the non-mapped earcon had an average of 5.6 (SD = 5.4) encounters with the earcon, while the 7 participants who selected the formant-based earcon had an average of 16.8 (SD = 6.2) encounters. Since the non-mapped earcon for `#spring` had a higher `f0` than the formant-based, this indicates that when making their selection, participants that had fewer encounters focused more on the emotional qualities of the earcon, while the participants with more encounters generally recognized the earcon they had encountered and were familiar with. In contrast, when analyzing the data of the `#joebiden` earcon, that saw a total of 204 encounters, only two participants selected the non-mapped earcon, and those participants had an average of 0.5 (SD=0.7) encounters, whereas the 6 participants who selected the formant-based earcon had an average of 9 (SD=7.5) encounters. Interestingly, the 11 participants who chose the mapped earcon without formants had an average of 13.4 (SD=6.6) encounters. This could indicate that either the formants might not have been present to the degree necessary in the earcons, or that the conditioning of melody is what dominates when reaching higher numbers of encounters. When comparing the frequency spectra of the earcons produced for `#joebiden` with formant synthesis and the one without the changes caused by the amplification of the formant frequencies are visually noticeable, however not distinctly different from the earcon without formant frequencies, which could speak for the

[7] Link to audio analysis software WaveSurfer: `https://sourceforge.net/projects/wavesurfer/`

theory of the formants not being clearly audible.

The restrictions caused by the current pandemic made it necessary to develop a method that would be suitable for digital testing, and the use of a web application as the training tool was chosen. This caused the amount of encounters each participant had with any earcon vary between participants, which made the results harder to interpret. Nevertheless, this resulted in a scenario that lets us examine how real users would react and respond to earcons of tweets arriving in real-time and gives some indication of if implicit learning did occur.

The fact that we could not control factors such as the playback device used by participants, could have had an effect on the presence of reproduction quality of the amplified formant frequencies. This is likely to be the case with small speakers, such as in laptops and mobile phones. There is also some degradation of audio quality induced by transmitting the earcon over Zoom during test sessions. One way of avoiding this problem in the future, would be to use timbres with broader spectrum such as those of cello or bassoon, or even multiple instruments at the same time. This would make amplified formant frequencies more prevalent. In fact, as shown in the appendix [3], the frequency spectra of both the mapped earcons with and without format synthesis appear quite similar, with flute being the timbre used.

6. CONCLUSIONS

In this study we have shown that users could identify Twitter hastags presented as earcons and that their recognition seems to be based on the conditioning of melodies used in the test, rather than the application of formant synthesis. However, an increase in the number of identified formant-based hashtags/earcons was observed corresponding to an increased number of times the users were exposed to them, indicating that a longer period of training can be of benefit; more encounters with formant-based earcons lead to higher rates of recognition.

In a recent test with 9 music experts we tested their association of formant-based earcons to different hashtags. We played for them some earcons they had never heard before via Zoom, and asked "Which word do you hear in the sound?". We provided two word alternatives per earcon. Answers were anonymously collected online: 7 participants recognized the `#joebiden` earcon (the choice was between `#joebiden` and `#donaldtrump`), 6 recognized the `#facebook` earcon (between `#facebook` and `#coffee`), and 5 recognized the `#coffee` one (between `#facebook` and `#coffee`). These are encouraging results that confirm the potential of the idea presented in this study.

7. ACKNOWLEDGEMENTS

The work presented in this paper was partially supported by a grant from NAVET center at the KTH Royal Institute of Technology [8].

[8] NAVET - A hub to navigate unexplored regions between art, technology and design, kth.se/navet

8. REFERENCES

[1] S. Brewster, P. Wright, and A. Edwards, "Experimentally derived guidelines for the creation of earcons," 01 1995.

[2] T. Hermann, A. Hunt, and J. Neuhoff, *The Sonification Handbook*, 01 2011.

[3] N. Fitz, K. Kushlev, R. Jagannathan, T. Lewis, D. Paliwal, and D. Ariely, "Batching smartphone notifications can improve well-being," *Computers in Human Behavior*, vol. 101, pp. 84–94, 2019.

[4] S. Bakker, E. van den Hoven, and B. Eggen, "Knowing by ear: Leveraging human attention abilities in interaction design," *Langmuir*, vol. 5, pp. 1–13, 01 2011.

[5] M. Engwall, A. Jerbrant, B. Karlsson, and P. Storm, *Modern Industrial Management*. Studentlitteratur AB, 2017.

[6] K. Kushlev, J. Proulx, and E. Dunn, ""Silence your phones": Smartphone notifications increase inattention and hyperactivity symptoms," 05 2016.

[7] D. A. Sumikawa, "Guidelines for the integration of audio cues into computer user interfaces," 6 1985.

[8] R. E. Remez, P. E. Rubin, D. B. Pisoni, and T. D. Carrell, "Speech perception without traditional speech cues," *Science*, vol. 212, no. 4497, pp. 947–950, 1981.

[9] H. Tankovska. Countries with the most Twitter users, 2021. [Online]. Available: https://tinyurl.com/49ebxasd

[10] H.-C. Chang, "A new perspective on twitter hashtag use: Diffusion of innovation theory," *Proceedings of the American Society for Information Science and Technology*, vol. 47, pp. 1 – 4, 11 2010.

[11] K. Giaxoglou, "#JeSuisCharlie? Hashtags as narrative resources in contexts of ecstatic sharing," *Discourse, Context & Media*, vol. 22, 09 2017.

[12] C. Gustafsson, "Sonic branding: A consumer-oriented literature review," *The Journal of Brand Management*, vol. 22, 01 2015.

[13] B. Walker and G. Kramer, "Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making," 2004.

[14] J. Sundberg, *The Science of the Singing Voice*. Northern Illinois University Press, 1987.

[15] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, vol. 24, pp. 175–184, 1951.

[16] P. Isotalo. Interactive IPA Chart. [Online]. Available: https://www.ipachart.com/

[17] J. Villalba and E. Lleida, "Detecting replay attacks from far-field recordings on speaker verification systems," 03 2011, pp. 274–285.