http://is.ijs.si

IS 2020

Zbornik 23. mednarodne multikonference

INFORMACIJSKA DRUŽBA

Zvezek H

Proceedings of the 23rd International Multiconference

INFORMATION SOCIETY

Volume H

Interakcija človek-računalnik v informacijski družbi

Human-Computer Interaction in Information Society

Uredili / Edited by

Veljko Pejović, Matjaž Kljun, Vida Groznik, Domen Šoberl, Klen Čopič Pucihar, Bojan Blažica, Jure Žabkar, Matevž Pesek, Jože Guna, Simon Kolmanič

7. oktober 2020 / 7 October 2020

Ljubljana, Slovenia

Zbornik 23. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2020

Zvezek H

Proceedings of the 23rd International Multiconference

INFORMATION SOCIETY – IS 2020

Volume H

Interakcija človek-računalnik v informacijski družbi
Human-Computer Interaction in Information Society

Uredili / Edited by

Veljko Pejović, Matjaž Kljun, Vida Groznik, Domen Šoberl, Klen Čopič Pucihar, Bojan Blažica, Jure Žabkar, Matevž Pesek, Jože Guna, Simon Kolmanič

http://is.ijs.si

7. oktober 2020 / 7 October 2020
Ljubljana, Slovenia

Uredniki:

Veljko Pejović
Univerza v Ljubljani, FRI, ACM SIGCHI Chapter Bled Slovenia

Matjaž Kljun
Univerza na Primorskem, FAMNIT

Vida Groznik
Univerza na Primorskem, FAMNIT

Domen Šoberl
Univerza na Primorskem, FAMNIT

Klen Čopič Pucihar
Univerza na Primorskem, FAMNIT

Bojan Blažica
Inštitut Jožef Stefan

Jure Žabkar
Univerza v Ljubljani, FRI

Matevž Pesek
Univerza v Ljubljani, FRI

Jože Guna
Univerza v Ljubljani, FE

Simon Kolmanič
Univerza v Mariboru, FERI

Založnik: Institut »Jožef Stefan«, Ljubljana
Priprava zbornika: Mitja Lasič, Vesna Lasič, Lana Zemljak
Oblikovanje naslovnice: Vesna Lasič

Dostop do e-publikacije:
http://library.ijs.si/Stacks/Proceedings/InformationSociety

Ljubljana, oktober 2020

Informacijska družba
ISSN 2630-371X

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=33206787
ISBN 978-961-264-200-6 (epub)
ISBN 978-961-264-201-3 (pdf)

PREDGOVOR MULTIKONFERENCI INFORMACIJSKA DRUŽBA 2020

Triindvajseta multikonferenca Informacijska družba (http://is.ijs.si) je doživela polovično zmanjšanje zaradi korone. Zahvala za preživetje gre tistim predsednikom konferenc, ki so se kljub prvi pandemiji modernega sveta pogumno odločili, da bodo izpeljali konferenco na svojem področju.

Korona pa skoraj v ničemer ni omejila neverjetne rasti IKTja, informacijske družbe, umetne inteligence in znanosti nasploh, ampak nasprotno – kar naenkrat je bilo večino aktivnosti potrebno opraviti elektronsko in IKT so dokazale, da je elektronsko marsikdaj celo bolje kot fizično. Po drugi strani pa se je pospešil razpad družbenih vrednot, zaupanje v znanost in razvoj. Celo Flynnov učinek – merjenje IQ na svetovni populaciji – kaže, da ljudje ne postajajo čedalje bolj pametni. Nasprotno - čedalje več ljudi verjame, da je Zemlja ploščata, da bo cepivo za korono škodljivo, ali da je korona škodljiva kot navadna gripa (v resnici je desetkrat bolj). Razkorak med rastočim znanjem in vraževerjem se povečuje.

Letos smo v multikonferenco povezali osem odličnih neodvisnih konferenc. Zajema okoli 160 večinoma spletnih predstavitev, povzetkov in referatov v okviru samostojnih konferenc in delavnic in 300 obiskovalcev. Prireditev bodo spremljale okrogle mize in razprave ter posebni dogodki, kot je svečana podelitev nagrad – seveda večinoma preko spleta. Izbrani prispevki bodo izšli tudi v posebni številki revije Informatica (http://www.informatica.si/), ki se ponaša s 44-letno tradicijo odlične znanstvene revije.

Multikonferenco Informacijska družba 2020 sestavljajo naslednje samostojne konference:

- Etika in stroka
- Interakcija človek računalnik v informacijski družbi
- Izkopavanje znanja in podatkovna skladišča
- Kognitivna znanost
- Ljudje in okolje
- Mednarodna konferenca o prenosu tehnologij
- Slovenska konferenca o umetni inteligenci
- Vzgoja in izobraževanje v informacijski družbi

Soorganizatorji in podporniki konference so različne raziskovalne institucije in združenja, med njimi tudi ACM Slovenija, SLAIS, DKZ in druga slovenska nacionalna akademija, Inženirska akademija Slovenije (IAS). V imenu organizatorjev konference se zahvaljujemo združenjem in institucijam, še posebej pa udeležencem za njihove dragocene prispevke in priložnost, da z nami delijo svoje izkušnje o informacijski družbi. Zahvaljujemo se tudi recenzentom za njihovo pomoč pri recenziranju.

V 2020 bomo petnajstič podelili nagrado za življenjske dosežke v čast Donalda Michieja in Alana Turinga. Nagrado Michie-Turing za izjemen življenjski prispevek k razvoju in promociji informacijske družbe je prejela prof. dr. Lidija Zadnik Stirn. Priznanje za dosežek leta pripada Programskemu svetu tekmovanja ACM Bober. Podeljujemo tudi nagradi »informacijska limona« in »informacijska jagoda« za najbolj (ne)uspešne poteze v zvezi z informacijsko družbo. Limono je prejela »Neodzivnost pri razvoju elektronskega zdravstvenega kartona«, jagodo pa Laboratorij za bioinformatiko, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani. Čestitke nagrajencem!

Mojca Ciglarič, predsednik programskega odbora
Matjaž Gams, predsednik organizacijskega odbora

FOREWORD
INFORMATION SOCIETY 2020

The 23rd Information Society Multiconference (http://is.ijs.si) was halved due to COVID-19. The multiconference survived due to the conference presidents that bravely decided to continue with their conference despite the first pandemics in the modern era.

The COVID-19 pandemics did not decrease the growth of ICT, information society, artificial intelligence and science overall, quite on the contrary – suddenly most of the activities had to be performed by ICT and often it was more efficient than in the old physical way. But COVID-19 did increase downfall of societal norms, trust in science and progress. Even the Flynn effect – measuring IQ all over the world – indicates that an average Earthling is becoming less smart and knowledgeable. Contrary to general belief of scientists, the number of people believing that the Earth is flat is growing. Large number of people are weary of the COVID-19 vaccine and consider the COVID-19 consequences to be similar to that of a common flu dispute empirically observed to be ten times worst.

The Multiconference is running parallel sessions with around 160 presentations of scientific papers at twelve conferences, many round tables, workshops and award ceremonies, and 300 attendees. Selected papers will be published in the Informatica journal with its 44-years tradition of excellent research publishing.

The Information Society 2020 Multiconference consists of the following conferences:

- Cognitive Science
- Data Mining and Data Warehouses
- Education in Information Society
- Human-Computer Interaction in Information Society
- International Technology Transfer Conference
- People and Environment
- Professional Ethics
- Slovenian Conference on Artificial Intelligence

The Multiconference is co-organized and supported by several major research institutions and societies, among them ACM Slovenia, i.e. the Slovenian chapter of the ACM, SLAIS, DKZ and the second national engineering academy, the Slovenian Engineering Academy. In the name of the conference organizers, we thank all the societies and institutions, and particularly all the participants for their valuable contribution and their interest in this event, and the reviewers for their thorough reviews.

For the fifteenth year, the award for life-long outstanding contributions will be presented in memory of Donald Michie and Alan Turing. The Michie-Turing award was given to Prof. Dr. Lidija Zadnik Stirn for her life-long outstanding contribution to the development and promotion of information society in our country. In addition, a recognition for current achievements was awarded to the Program Council of the competition ACM Bober. The information lemon goes to the "Unresponsiveness in the development of the electronic health record", and the information strawberry to the Bioinformatics Laboratory, Faculty of Computer and Information Science, University of Ljubljana. Congratulations!

Mojca Ciglarič, Programme Committee Chair
Matjaž Gams, Organizing Committee Chair

KONFERENČNI ODBORI
CONFERENCE COMMITTEES

International Programme Committee

Vladimir Bajic, South Africa
Heiner Benking, Germany
Se Woo Cheon, South Korea
Howie Firth, UK
Olga Fomichova, Russia
Vladimir Fomichov, Russia
Vesna Hljuz Dobric, Croatia
Alfred Inselberg, Israel
Jay Liebowitz, USA
Huan Liu, Singapore
Henz Martin, Germany
Marcin Paprzycki, USA
Claude Sammut, Australia
Jiri Wiedermann, Czech Republic
Xindong Wu, USA
Yiming Ye, USA
Ning Zhong, USA
Wray Buntine, Australia
Bezalel Gavish, USA
Gal A. Kaminka, Israel
Mike Bain, Australia
Michela Milano, Italy
Derong Liu, Chicago, USA
prof. Toby Walsh, Australia

Organizing Committee

Matjaž Gams, chair
Mitja Luštrek
Lana Zemljak
Vesna Koricki
Marjetka Šprah
Mitja Lasič
Blaž Mahnič
Jani Bizjak
Tine Kolenik

Programme Committee

Mojca Ciglarič, chair
Bojan Orel, co-chair
Franc Solina,
Viljan Mahnič,
Cene Bavec,
Tomaž Kalin,
Jozsef Györkös,
Tadej Bajd
Jaroslav Berce
Mojca Bernik
Marko Bohanec
Ivan Bratko
Andrej Brodnik
Dušan Caf
Saša Divjak
Tomaž Erjavec
Bogdan Filipič
Andrej Gams
Matjaž Gams
Mitja Luštrek
Marko Grobelnik
Nikola Guid
Marjan Heričko
Borka Jerman Blažič Džonova
Gorazd Kandus
Urban Kordeš
Marjan Krisper
Andrej Kuščer
Jadran Lenarčič
Borut Likar
Janez Malačič
Olga Markič
Dunja Mladenič
Franc Novak
Vladislav Rajkovič
Grega Repovš
Ivan Rozman
Niko Schlamberger
Špela Stres
Stanko Strmčnik
Jurij Šilc
Jurij Tasič
Denis Trček
Andrej Ule
Tanja Urbančič
Boštjan Vilfan
Baldomir Zajc
Blaž Zupan
Boris Žemva
Leon Žlajpah

KAZALO / TABLE OF CONTENTS

Zbornik 23. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2020

Zvezek H

Proceedings of the 23rd International Multiconference

INFORMATION SOCIETY – IS 2020

Volume H

Interakcija človek-računalnik v informacijski družbi
Human-Computer Interaction in Information Society

Uredili / Edited by

Veljko Pejović, Matjaž Kljun, Vida Groznik, Domen Šoberl, Klen Čopič Pucihar, Bojan Blažica, Jure Žabkar, Matevž Pesek, Jože Guna, Simon Kolmanič

http://is.ijs.si

7. oktober 2020 / 7 October 2020
Ljubljana, Slovenia

PREDGOVOR

Interakcija človek–računalnik v informacijski družbi je konferenca, ki jo organizira Slovenska skupnost za
proučevanje interakcije človek–računalnik. Namen konference je zbrati raziskovalce, strokovne delavce in študente s področja in ponuditi možnost izmenjave izkušenj in raziskovalnih rezultatov, kakor tudi navezave stikov za bodoča sodelovanja .

Tokratna, peta reinkarnacija konference se že drugič odvija pod okriljem SIGCHI poglavja ACM Chapter Bled, ki je nastalo tudi kot posledica prejšnjih konferenc. O rasti HCI skupnosti v regiji pa priča tudi vse večje število prispevkov, ki prihajajo z vseh večjih visokošolskih zavodov v Sloveniji.

Teme, ki jih konferenca pokriva segajo od bolj uveljavljenih, kot so vizualizacija, snovanje grafičnih in uporabniških vmesnikov, ki temeljijo na govoru, personalizacija in prilagajanje interakcije uporabnikom, pa do virtualne in nadgrajene resničnosti ter uporabniških vmesnikih v turizmu, umetnosti in e-učenju.

FOREWORD

Human-computer interaction in information society is a conference organized by the Slovenian HCI community. The purpose of the conference is to gather researchers, practitioners and students in the field and offer the opportunity to exchange experiences and research results, as well as to establish contacts for future cooperations.

This year's fifth reincarnation of the conference is, for the second time, organized by the SIGCHI Chapter ACM Chapter Bled, which has been established also as a result of previous conferences. The growth of the HCI community in the region is witnessed by the doubled number of contributions coming from all major higher education institutions in Slovenia.

The topics covered by the conference range from the more established ones, such as visualization and design of graphical and audio user interfaces, personalisation and interaction adaptation, to virtual and augmented reality, and the ap plication of user interfaces in tourism, arts, and e-learning.

PROGRAMSKI ODBOR / PROGRAMME COMMITTEE

Nuwan T Attygalle (Univerza na Primorskem)

Bojan Blažica (Inštitut Jožef Stefan)

Klen Čopič Pucihar (Univerza na Primorskem)

Jordan Deja (Univerza na Primorskem)

Vida Groznik (Univerza na Primorskem)

Jože Guna (Univerza v Ljubljani)

Matjaž Kljun (Univerza na Primorskem)

Simon Kolmanič (Univerza v Mariboru)

Ines Kožuh (Univerza v Mariboru)

Elham Motamedi (Univerza na Primorskem)

Marko Tkalčič (Univerza na Primorskem)

Domen Šoberl (Univerza na Primorskem)

Veljko Pejović (Univerza v Ljubljani)

Jure Žabkar (Univerza v Ljubljani)

Investigating the Role of Context and Personality in Mobile Advertising

Andrej Martinovič
Faculty of Computer and Information Science, University of Ljubljana, Slovenia
am6694@student.uni-lj.si

Veljko Pejović
Faculty of Computer and Information Science, University of Ljubljana, Slovenia
Veljko.Pejovic@fri.uni-lj.si

ABSTRACT

More than three billion smartphones carried by their users at virtually all times, represent an unprecedented platform for in-situ advertisement delivery. While recent efforts in data analysis and machine learning led to significant advances in the way relevant content is selected to be shown to a user, thorough investigation on how the content should be displayed to a mobile user is yet to be conducted. In this work we present our preliminary research on the role of the context in which an advertisement is consumed and the personality of a user consuming it on the perception of the ad content. We conduct a 7-week study with 14 mobile users who were exposed to both video and picture ads. Through mobile sensing and experience sampling we capture the information on the context in which the ad was seen, the user's attitude towards the ad, as well as the user's personality traits. Statistical analysis based on mixed-effect modelling demonstrates that personality traits play an important role in ad perception, as does the ad type, with picture ads being preferred to video ads, while the effect of the context on ad perception appears to be negligible.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Ubiquitous and mobile devices**; Empirical studies in ubiquitous and mobile computing.

KEYWORDS

mobile advertising, multilevel models, ubiquitous computing

1 INTRODUCTION AND BACKGROUND

Tremendous amounts of digital traces, just-in-time sensor information, and the advances in data processing have resulted in major shifts in how the advertising is performed. Machine learning and recommender systems are at the core of modern advertising solutions [9]. The selection of the ad to be show to the user benefits from the history of purchases, information on the similarity among users, but also on the information about a user's personality [6].

Moving to the mobile domain, contextual information, such as location may impact the relevance of an ad [2]. The context, that can be sensed by a smartphone, such as a user's location, his physical activity, time of day, and other factors, can also be used to determine the suitability of a moment for information delivery [7].

While the previous work focuses on the content or the timing of the ad delivery, the type of the ad, to the best of our knowledge, has not been examined in the mobile domain. Nevertheless, the type of the ad, whether it is a picture, a short or a long video, or perhaps an interactive content (e.g. a short game) is an important parameter that influences the overall design of an ad, the platforms at which the ad can be shown, advertisement budget, etc. In this paper we focus on the perception of an ad type in mobile computing and pose the following research question: *Can the contextual information collected by the mobile phone sensors and the information on a user's personality predict a user's perception of different types of mobile ads?*

2 METHODOLOGY

To obtain ecologically valid data on mobile ad perception in different contexts we developed a data collection mobile application that serves ads, captures a user's attitudes towards the displayed ads, and collects sensor data pertaining to the context of use. In the rest of the section we present the details of our app.

Mobile Application

We implemented a full-fledged mobile app that caters to the need of our target users – students at our University. The application was built for the Android platform and serves as a utility tool allowing its users to: obtain information on nearby restaurants providing subsidised student meals, get real-time public transport timetables, record or share important student notes, retrieve latest student related news

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

Figure 1: Data collection app: one of the functionalities (left), advertisement (center) and an ESM questionnaire (right).

feeds, save and access their most needed school gadgets, and organise their class schedules (Figure 1 left).

Mobile ads. Mobile ads come in different flavours ranging from simple picture-based ads, over video ads, to more interactive game-like ads. We opted to investigate the two most frequent types of ads in our study – pictures and videos. We further divide the video ads into two groups – short videos, with the length of 30 seconds or less, and long videos with the length between 30 and 80 seconds. From each of the three groups – pictures, short videos, and long videos – we gathered 31 different publicly available ads and pre-loaded them on our server. After five actions that a user makes within our app, a request is made to our back-end system which responds with a random ad of a randomly chosen category. Simultaneously, we activate mobile phone's sensors and capture the user's context, including the physical activity (through Android's Google Activity Recognition functionality), location (clustered as *work*, *home*, or *other*, according to the method described in [7]), screen brightness, battery level, time of day, and the Internet connectivity type.

Experience sampling method (ESM) questionnaires. ESM is commonly used to gather the participants own thoughts, emotions, behaviour, etc [3]. In our case it provided us with feedback regarding the participants assessment of overall ad suitability. With the included questionnaire we also wanted to measure the interaction level between the user and the displayed ad. Thus, the questionnaire consisted of following questions: what was shown on the ad, which brand/trademark was advertised, and was the ad shown in an appropriate form. The first two questions were used to assess whether the user was engaged with the ad. The last question focused on the appropriateness of the displayed ad. The answers are recorded with five-level Likert scales. Figure 1 represents the data collecting workflow, where a user made an action, which led to the ad being displayed, followed by the ESM questionnaire.

Personality test. Previous research demonstrates that personality traits have a moderate effect on a user's attitude towards advertisement [1]. Therefore, we included the BFI-10 personality test [8] as a part of our app. The test includes ten questions about a user's traits answered on a seven-point Likert scale. The processed BFI-10 data, assessing a user's personality along the five dimensions (extraversion, agreeableness, openness, conscientiousness, and neuroticism), was further compared to the statistics calculated on a larger population set in order to extract the percentiles to which the participants personality trait scores belong [8].

Data collection campaign

Our data collection campaign lasted for seven weeks in spring 2020 and included 14 participants who in total viewed 994 ads, out of which 501 they labeled, i.e. an ESM questionnaire was completed immediately after the ad was viewed. The distribution of labeled and unlabeled ad types is roughly even. The viewing was reasonably evenly distributed among users, with the least active participant contributing 2.4% and the most active participant contribution 12.6% of the data. In our study we included 12 picture ads, 9 short video ads, and 10 long video ads. The ads were randomly shown both within and among users, i.e. each two users saw different ads where a participant shown a picture ad from a specific brand need not have seen a video ad from the same trademark. The majority of viewed ads, that are labeled, were pictures (40.5%), followed by short videos (34.7%). The least amount of user feedback was from long videos (24.8%). The average score (questionnaire answers ranging from "Strongly disagree" to "Strongly agree" were transformed to the integer [-2, 2] scale) over all ads was 0.377, yet it differs across the ad types. Labeled pictures had an average score of 0.695, short videos 0.253, and long videos 0.032.

3 MOBILE AD PERCEPTION MODELLING

Our data collection study elaborated in Section 2 has resulted in a heterogeneous dataset with an uneven number of datapoints across users, across contextual characteristics, and ad types. The natural organisation of our data into groups makes multilevel modelling-based analysis particularly appropriate. Such models generalise the linear regression in a manner that allows that the effect of a group (e.g. a particular user, a personality type, etc.) is disentangled from the effect of predictors, such as contextual variables [4] [5].

With hierarchical modeling we gradually increase the model complexity by including different parameters as a part of fixed or random effects. At each step we need to compare our new model to the previous one. This is done by preforming a chi-squared test checking if the residual sum of squares of the new model is statistically significantly smaller than that of the old model. To further verify which model is better we calculated the AIC (Akaike information criterion) and BIC (Bayesian Information Criterion) metrics, where smaller values indicate a better model, since the relative amount of information lost is lower.

In this section we present the results of multilevel modelling with two models constructed on the labeled data in order to investigate the impact of different parameters on the ad perception – a model where the *user ID* is the grouping variable and a model where the user's *personality* is the grouping variable. We then use both labeled and unlabeled data in a semi-supervised learning fashion to construct our final predictive model rooted in users' personalities.

User ID-based model

The basic user model includes merely the participants' IDs as the grouping variable. From there on we gradually increase the model complexity by separately adding context-based parameters. We experiment with the inclusion of the physical activity, location, screen brightness, battery level, time of day, and the internet connectivity type information in our model, and find that none of the contextual variables have a statistically significant influence on whether a user marks an ad as appropriate or not. In addition, the comparison of the basic model with the context-based ones reveals that the AIC and BIC metrics increase, and the p-value of the chi-squared model comparison remains above the 0.05 threshold, again indicating the superiority of the basic model.

Since the context is shown to be irrelevant, we focus on the content and the type-based models. With the inclusion of ad type, as a part of fixed effects, we were able to build a model that preforms better then the basic one. We suspect that different users score different ad types in different manners, thus we included the type parameter as a random slope. Metrics AIC, BIC show a significant decrease, indicating that the new model preforms better than the previous one. The analysis of the model reveals that picture ads receive a predominantly positive score, short videos neutral-negative, and long videos very negative score. Slope coefficients for ad type were also found to be varying within users. We further experiment with content-based models, where the each particular ad is encoded as its own content category.The AIC, BIC, and chi-square-based comparison indicate that the content has a statistically significant impact on ad scoring. With both content and ad type being relevant we further investigate whether it is possible to combine both models and also include the ad viewing duration as a parameter. Indeed, our best preforming model includes *the duration of ad watching*, and cross-level *interaction of ad content and ad type* as fixed effects, and *ad type* as the random effect. The conditional R^2 metric of such a model is 0.455 whilst the marginal R^2 is 0.204 indicating a reasonably good fit.

Personality-based model

The above user ID-based model demonstrates the impact of individual traits on the ad perception. Nevertheless, the model is not suitable for real-world use, as it requires that an individual's data is available *before* predictions can be made. Therefore, we now design a model that, instead of data from a particular user, is based on the information about personality traits of a user. Such information can be obtained quickly through a personality test.

The basic personality-based model only includes a grouping variable based on personality traits without any fixed effects or random slopes. As before, we find that the inclusion of context parameters does not improve the basic model so we focus on the ad content and ad type as the next modeling level. Gradually increasing the complexity of our model we come to similar conclusions as in the previous section. The fixed effects include a cross-level interaction of ad content and ad type, where the random effects include ad type only. The final personality-based model demonstrates that ad types are marked differently within different personality groups. One particular group consisting of extrovert, nonconflicting, non-conscious, and emotionally stable users is found to stand out. In the mentioned group pictures had an average score of -0.4, short videos 0.636 and long videos -0.75. To see if the scores were indeed significantly different, we preform a Welch's t-test between this outlying and all other personality groups (Table 1). We find that the difference in short video scoring between the compared groups is not statistically significant, whilst the scores of pictures are.

Metrics	Pictures	Short videos	Long videos
t-test	-4.087	1.026	-1.545
p-value	0.001	0.326	0.162
95% conf. interval	[-1.771, -0.565]	[-0.467, 1.286]	[-2.089, 0.416]
Outlying group avg.	-0.4	0.636	-0.75
Other groups avg.	0.768	0.227	0.086

Table 1: Welch's t-test between the outlying personality group (extrovert, non-conflicting, non-conscious, and emotionally stable) and other personality groups.

Even though we built a personality-based model with the intent to make it more general, we found that not all personality combinations are included, as our sample size is not large enough. With 14 participants, out of 16 different possible personality groups (openness omitted) only 7 are covered. The final model's R^2 metric conditional value is 0.377 and the marginal is 0.198.

Predictive personality-based model

The user ID-based model demonstrates that *who* is watching the ad is more important than *in what situation* is someone watching the ad. Predictions of an attitude towards an ad could be used to decide whether to show an ad of a certain type, or whether to show an ad at all. Yet, personalised user-based models would require labeled data for each user, making their construction impractical. The analysis of the personality-based multilevel models demonstrates that general personality traits, obtainable through a simple 10-item questionnaire, can be used to build an informative model. Here we examine the predictive potential of a fully generaliseable model based on personality traits information.

With semi-supervised learning, we first label the unlabeled data – using the previously constructed user ID-based model, we predict the labels for the 493 unlabeled points. We then proceed with constructing a new personality-based model. Repeating the gradual increase of complexity procedure we find that the following context variables significantly impact the fit: screen brightness, battery level, and Internet connection type. Nevertheless, the variables do not feature highly in the final model, as ad content and ad type prove to be much more impactfull on the final ad scoring. Our final generalised personality-based model constructed on all gathered data includes a cross-level interaction of ad content and ad type as fixed effects and ad type as a random effect.

To assess the potential of the model to correctly predict the score a previously unseen user will give to an ad in a certain situation, we perform a leave-one-person out evaluation and in each step calculate the (root) mean square error (R)MSE and mean absolute error (MAE) of our model and the baseline model that predicts the mean score across the dataset. Average RMSE, MSE, and MAE for the personality-based model are 0.967, 1.014, and 0.785, whereas the baseline results in 1.117, 1.347, and 0.865, respectively, indicating that the personality-based predictive model fits the data better than the majority classifier. The R^2 metric's conditional value of the model is 0.488 and the marginal is 0.308.

4 DISCUSSION AND CONCLUSION

In this paper we examined of the role of context and a user's personality on ad perception. While our initial assumption was that users would prefer either picture or video ads depending on the context of viewing, we discovered that picture ads are almost universally better accepted. This surprising finding might stem from our data collection limitations – conducted during the COVID-19 pandemics, the data fails to capture the full range of locations and activities we would expect to see during regular times. A prominent role of a user's personality in the perception of an ad is another interesting finding. We discover that certain personalities actually prefer short videos over picture ads. Our general predictive model takes personalities into account and is able to predict the attitude that a previously unobserved user will have towards an ad better than the baseline model. The initial analysis also demonstrates that the content of the ad, a property that was outside of the scope of our study, may significantly impact the perception and should be further examined.

REFERENCES

[1] Aliosha Alexandrov, Susan Mayers, and Sandipan Sen. 2010. The moderating effect of personality traits on attitudes toward advertisements: A contingency framework. *Management Marketing* 5 (01 2010).

[2] Christine Bauer and Christine Strauss. 2016. Location-based advertising on mobile devices: A literature review and analysis. *Management Review Quarterly* 66 (01 2016), 159–194.

[3] Niall Bolger. 2013. *Intensive longitudinal methods: An introduction to diary and experience sampling research.* Guilford Press.

[4] Andrea Feldstain, Heather Woltman, Jennifer MacKay, and Meredith Rocci. 2012. Introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology* 8 (02 2012), 62–69.

[5] Andrew Gelman and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

[6] Haris Kriještorac, Rajiv Garg, and Maytal Saar-Tsechansky. 2019. Personality-Based Content Engineering for Rich Digital Media. In *AIS BLED*. Bled, Slovenia.

[7] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. In *ACM UbiComp*. Seattle, WA, USA.

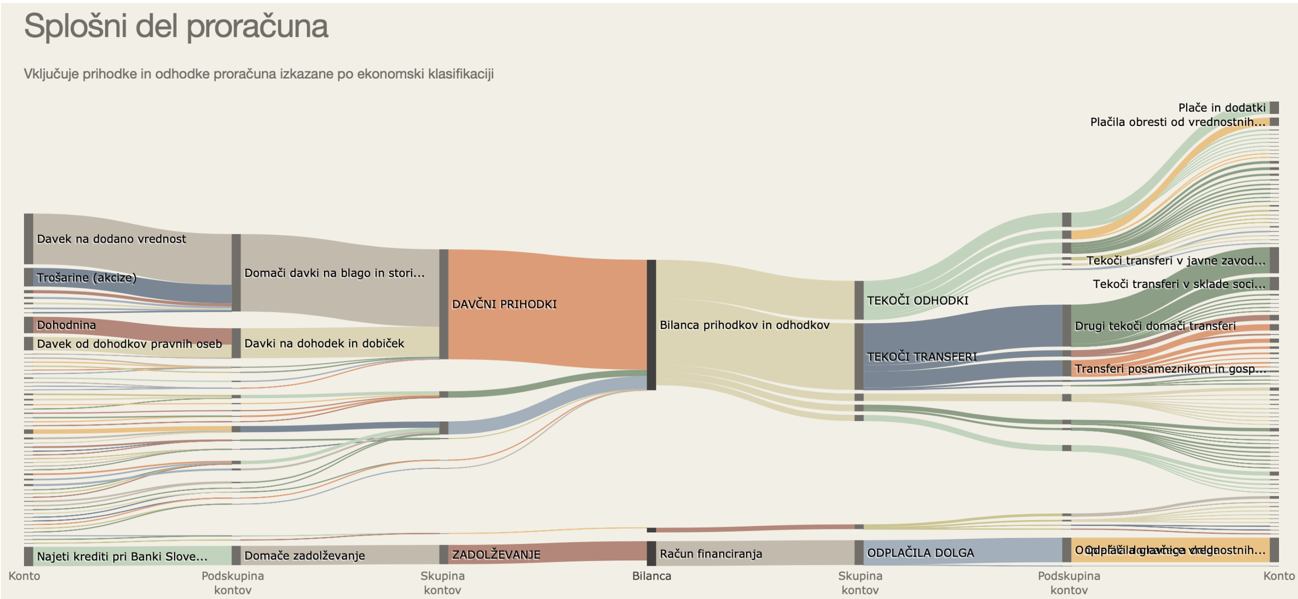
[8] Beatrice Rammstedt and Oliver P John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41, 1 (2007), 203 – 212.

[9] Ayush Singhal, Pradeep Sinha, and Rakesh Pant. 2017. Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works. *Intl. Journal of Computer Applications* 180, 7 (Dec 2017), 17–22.

Interaktivna vizualizacija proračuna Republike Slovenije s Sankeyevim diagramom

Interactive Visualization of the Slovenian Budget with the Sankey Diagram

Tea Tušar
tea.tusar@ijs.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia



Slika 1: Sankeyev diagram za splošni del proračuna za leto 2020

POVZETEK

Predstavljamo spletno aplikacijo z interaktivno vizualizacijo proračuna Republike Slovenije. Z dvema Sankeyevima diagramoma, ki prikazujeta različne kategorije proračuna in denarne tokove med njimi, vizualiziramo tako splošni kot posebni del državnega proračuna. Interakcija omogoča spreminjanje pogledov, s katerimi lahko prikažemo več podrobnosti. Aplikacija ne ponuja vnaprej izbranih vidikov proračuna, ampak je namenjena prostemu raziskovanju po njegovih podatkih in kot taka predstavlja alternativo obstoječim vizualizacijam proračuna. Na voljo je na naslovu http://proracun.herokuapp.com.

KLJUČNE BESEDE

državni proračun, interaktivna vizualizacija, Sankeyev diagram

ABSTRACT

We present a web application with interactive visualizations of the Slovenian budget. With two Sankey diagrams that show different budget categories and the cash flows between them, we visualize both the general and the specific budget part. Interaction allows to change views, so that more details can be shown. The application does not produce pre-selected aspects of the budget, but is intended for free searching among its data and as such represents an alternative to existing budget visualizations. It is available at http://proracun.herokuapp.com

KEYWORDS

state budget, interactive visualization, Sankey diagram

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

1 UVOD

Živimo v času velepodatkov, družabnih omrežij in takojšnje komunikacije, ki nam v vsakem trenutku nudijo ogromne količine informacij. Ta preobremenjenost z informacijami nam otežuje poglabljanje vanje in njihovo preverjanje. Tako se pogosto zanašamo na tuje interpretacije in se nehote znajdemo v pasivni vlogi prejemnika informacij, ki so lahko tudi pomanjkljive ali (namenoma) napačne.

Temu se je moč izogniti z lastnim preverjanjem podatkov, ki pa je lahko zelo zahtevno opravilo. Na voljo moramo imeti dostop do podatkov, možnost obdelave velike količine podatkov, metode za vizualizacijo ter znanje potrebno za umestitev podatkov v širši kontekst in njihovo pravilno interpretacijo. Pri tem nam lahko pomagajo orodja, ki pridobivanje in obdelavo podatkov opravijo namesto nas.

V prispevku predstavljamo novo takšno orodje (slika 1), na voljo na naslovu http://proracun.herokuapp.com, ki preko interaktivne vizualizacije s t.i. Sankeyevim diagramom uporabniku pomaga pri razumevanju proračuna Republike Slovenije in iskanju informacij v njem. To je relevantno, saj je državni proračun največja izmed štirih blagajn javnega financiranja in predstavlja nekaj manj kot polovico vseh javnofinančnih odhodkov [10]. Razkriva fiskalno politiko, razvojne cilje, prednostna področja ter politične in strateške prednostne naloge vlade.

Kot pri vseh vizualizacijah, ki predstavljajo kompleksne podatke, je tudi tu bistvenega pomena interaktivnost. Pri obsežnih zbirkah podatkov namreč zaradi omejitev ljudi na eni in računalniških vizualizacij na drugi strani ni mogoče vseh podatkov pokazati naenkrat. Bolje se izkaže interakcija, pri kateri uporabnik s svojimi dejanji sproža spremembe pogledov. Za razliko od statičnega pogleda, ki lahko naenkrat prikaže samo en vidik podatkov, interakcija podpira številne poizvedbe. Še posebej je koristna pri preiskovanju na več ravneh podrobnosti, ko nam omogoča, da se (postopoma) premaknemo od pregleda na najvišji ravni preko vmesnih pregledov do najbolj podrobnega pregleda, ki lahko prikazuje le majhen del vseh podatkov [5].

Naloga, ki jo novo orodje naslavlja, ni predstavljanje ali razlaganje vnaprej izbranih vidikov proračuna, temveč podpora pri prostem raziskovanju po njegovih podatkih, ki uporabniku pomaga, da najde lastne vpoglede vanje. Kot tako je torej dopolnitev obstoječim vizualizacijam proračuna, kot so razlagalne infografike in druge vizualizacije, ki jih pripravlja Ministrstvo za finance Republike Slovenije (več o njih v razdelku 2.3). Orodje je namenjeno tako navadnim državljanom kot novinarjem in drugim profilom, ki jih proračun tako ali drugače zadeva in ga želijo raziskati ter tako bolje razumeti.

V nadaljevanju najprej na kratko predstavimo državni proračun, njegovo strukturo, dostopnost podatkov in obstoječe vizualizacije. Nato se posvetimo novemu orodju za vizualizacijo proračuna. Po opisu Sankeyevega diagrama razložimo kako ga lahko obogatimo z uporabo interakcije. Predstavimo tudi podrobnosti izdelave vizualizacije in razpravljamo o njenih lastnostih. Prispevek zaključimo s povzetkom in zamislimi za nadgradnjo orodja.

2 DRŽAVNI PRORAČUN

Državni proračun Republike Slovenije je gospodarsko-politični akt, ki vključuje predvidene prihodke in odhodke države za eno leto. Sprejme ga Državni zbor po predpisanem postopku. Kadar so dejanski prihodki manjši od načrtovanih ali nastanejo nove obveznosti, ki v proračunu niso bile predvidene, vlada lahko predlaga *rebalans proračuna*[1]. Z njim proračun uskladi s spremenjenimi okoliščinami.

Državni proračun je ena od štirih blagajn javnega financiranja. Preostale tri so *pokojninska blagajna*, iz katere se pretežno izplačujejo pokojnine in invalidnine, *zdravstvena blagajna*, ki pokriva predvsem stroške delovanja zdravstvenih domov, bolnišnic in zdravil ter *občinski proračuni*, ki obsegajo prihodke in odhodke vseh 212 občin. Največja blagajna je ravno državni proračun, ki predstavlja 48,4 % vseh javnofinančnih odhodkov. Sledijo pokojninska blagajna s 27,1 % odhodkov, zdravstvena blagajna s 14,2 % odhodkov in občinski proračuni z 10,3 % odhodkov [10].

[1]V času pisanja tega prispevka se pripravlja rebalans proračuna za leto 2020 [13]. Povod zanj je izraziti upad proračunskih prihodkov med epidemijo COVID-19, hkrati pa rast izdatkov zaradi sprejetih ukrepov vlade za omilitev posledic krize in ohranitev gospodarske aktivnosti.

2.1 Struktura proračuna

Državni proračun je sestavljen iz treh delov.

- **I. del: Splošni del proračuna** vključuje bilanco prihodkov in odhodkov, račun finančnih terjatev in naložb ter račun financiranja. Izkazuje se po ekonomski klasifikaciji (skupina kontov, podskupina kontov in konto).
- **II. del: Posebni del proračuna** izkazuje porabo javnofinančnih sredstev posameznih proračunskih uporabnikov preko institucionalne klasifikacije (nadskupina proračunskih uporabnikov, skupina proračunskih uporabnikov in proračunski uporabnik) ter vključuje odhodke in druge izdatke delovanja predstavljene po programski klasifikaciji (politika, program in podprogram).
- **III. del Načrt razvojnih programov** predstavlja načrt odhodkov po podprogramih, ukrepih, skupinah projektov, projektih in virih financiranja po posameznih letih za celotno obdobje izvajanja projektov in ukrepov.

Tako za splošni kot za posebni del proračuna so na voljo tudi dodatne obrazložitve. V nadaljevanju se osredotočamo le na ta dva dela proračuna.

2.2 Dostopnost podatkov

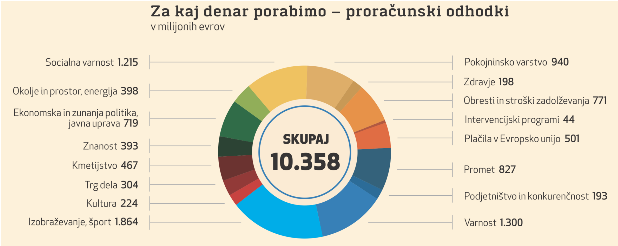
Na spletišču državne uprave (https://www.gov.si/) je pod okriljem Ministrstva za finance podanih mnogo informacij o državnem proračunu [8]. Med njimi so prosto dostopni tudi podatki o sprejetih proračunih za vsa leta med letoma 2004 in 2021. Ti so na voljo v tabelarični obliki v datotečnem formatu PDF za vse tri dele proračuna. Namenjeni so torej predvsem pregledu in niso primerni za dodatno računalniško obdelavo.

Ravno nadaljnji obdelavi pa so namenjeni podatki v datotečnem formatu CSV na portalu Odprti podatki Slovenije (OPSI, https://podatki.gov.si/). Na portalu sta za vse proračune med letoma 2014 in 2021 na voljo splošni in posebni del proračuna, od leta 2019 naprej pa še načrt razvojnih programov [11]. Vsi podatki uporabljeni v tem prispevku izhajajo iz portala OPSI.

2.3 Obstoječe vizualizacije

Ministrstvo za finance poleg golih podatkov o proračunu od leta 2017 naprej objavlja tudi infografike s ključnimi podatki o proračunu, s katerimi želi proračun približati širšemu krogu državljanov. Primer takšne infografike je prikazan na sliki 2 [9]. Infografika izpostavlja določene vidike posebnega proračuna – v tem primeru 16 politik, pri čemur so nekatere združene, saj je originalnih politik, zajetih v proračun, 24. Manjše število politik je lažje za razumevanje, a neizogibno skriva nekatere podrobnosti.

Poleg infografik so od začetka leta 2020 na voljo tudi tri vrste (interaktivnih) vizualizacij proračuna [12]. Prva omogoča vpogled v trenutno stanje prihodkov in odhodkov proračuna, ki se dnevno osvežuje. Iz nje je razvidno ali se proračun izvaja skladno s pričakovanji. Druga vizualizacija je interaktivna in za vseh 24 politik proračuna omogoča podrobnejši pogled porabe v posebnem oknu, v katerem so odhodki dodatno razdeljeni po programih in kontih. Odhodki vsake politike so prikazani tudi za pretekla leta (od leta 2009 naprej). Tretja vizualizacija pa nudi vpogled v posamezne projekte, kjer interaktivnost omogoča iskanje projektov po različnih kriterijih, med drugim tudi po tem v kateri regiji in občini se izvajajo.



Slika 2: Infografika bilance odhodkov za posebni proračun za leto 2020 (vir: Ministrstvo za finance [9])

3 INTERAKTIVNA VIZUALIZACIJA S SANKEYEVIM DIAGRAMOM

Kot dopolnitev obstoječim grafičnim prikazom predlagamo vizualizacijo proračuna z dvema Sankeyevima diagramoma – enim za splošni in drugim za posebni del proračuna.

3.1 Sankeyev diagram

Sankeyev diagram (poznan tudi kot aluvialni diagram) prikazuje kategorije in kvantitativne odnose med njimi [4]. Kategorije so vizualizirane kot pravokotniki (na sliki 1 obarvani v sivo), odnosi med njimi pa kot tokovi (na sliki 1 v različnih barvah). Širina toka je sorazmerna s količino, ki povezuje dve kategoriji.

Čeprav je Sankeyev diagram poimenovan po diagramih energetske učinkovitosti parnega stroja Matthewa Sankeya iz leta 1898 [3], je bil v rabi že prej. Eden najbolj poznanih Sankeyevih diagramov je Napoleonova ruska kampanija, ki jo leta 1869 ustvaril Charles Minard [7].

Sankeyev diagram je videti kot nalašč za vizualizacijo proračunskih podatkov, saj lahko na eni sami sliki prikaže mnogo raznolikih prihodkov in odhodkov ter morebitno razliko med njunima vsotama. Na sliki 1 so bilance označene s temno sivo barvo in postavljene na sredino grafičnega prikaza. Kategorije na levi kažejo prihodke v proračun, kategorije na desni pa njegove odhodke. Sankeyev diagram dobro prikaže tudi kako se neka kategorija razčleni na več podkategorij in kakšna so razmerja med njimi. Na sliki je to vidno za hierarhijo bilanca – skupina kontov – podskupina kontov – konto (na strani prihodkov in odhodkov).

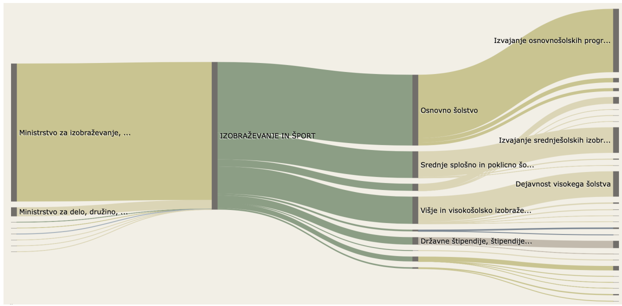
Sankeyevega diagrama za posebni del proračuna zaradi omejenega prostora v prispevku ne prikazujemo v celoti (v nadaljevanju bomo videli nekatere njegove dele).

3.2 Uporaba interakcije

Sankeyevemu diagramu lahko izrazno moč povečamo z uporabo interakcije. Predlagano orodje podpira naslednje interakcije:

- *Izpis več podatkov.* Ker se zneski v državnem proračunu med seboj precej razlikujejo, so nekatere kategorije in tokovi lahko zelo debeli, drugi pa komaj vidni. Poleg tega je na določenih ravneh število kategorij in tokov precejšnje. To pomeni, da ne moremo izpisati imena vseh kategorij, ampak se omejimo le na največje. Interakcijo lahko koristimo za to, da se imena kategorij (tudi tistih najmanjših) v celoti izpišejo šele takrat, ko se z miško postavimo nad njimi (glej sliko 3 zgoraj). Na podoben način interakcijo uporabimo tudi pri premikanju miške nad tokovi, kjer se ob tem pokaže več informacij o toku (njegov izvor in ponor ter znesek, slika 3 spodaj).

Slika 3: Izpis dodatnih podatkov ob interakciji s kategorijo (zgoraj) in tokom (spodaj) posebnega dela proračuna za leto 2020



Slika 4: Podrobnejši pregled kategorije Izobraževanje in šport posebnega dela proračuna za leto 2020

- *Sprememba pogleda.* S klikom na kategorijo spremenimo pogled tako, da se približamo izbrani kategoriji in vsem njenim podrejenim kategorijam (ter v primeru posebnega proračuna tudi njenim prvim nadrejenim kategorijam). Na ta način lahko prikažemo kategorije in tokove, ki so v izvirnem pogledu predrobni ali preveč nagneteni, da bi jih lahko dobro videli. Primer takšne spremembe pogleda na diagramu posebnega dela proračuna je ilustriran na sliki 4. Tu je kategorija Izobraževanja in športa povečana čez celoten zaslon, kar nam omogoča, da v podrobnosti vidimo njene podkategorije in njihova medsebojna razmerja. Hkrati pa vidimo tudi katera ministrstva so odgovorna za to politiko. V takšnem pogledu se lahko odločimo, da nadaljujemo s pregledovanjem drugih kategorij (s klikom nanje) ali pa se s klikom na katerikoli tok vrnemo na prvotni pogled.
- *Izbira podatkov.* Preko zavihka (ni viden na slikah) lahko izberemo leto proračuna, ki nas zanima. Trenutno imamo na voljo podatke za proračune za leta 2019, 2020 in 2021. Ob spremembi leta se izrišeta nova dva Sankeyeva diagrama (za splošni in posebni del proračuna), ki vsebujeta podatke za izbrano leto.

3.3 Izdelava vizualizacije

3.3.1 Priprava podatkov. Kot že omenjeno, so vsi podatki, uporabljeni v tem orodju, pridobljeni s portala OPSI [11]. Podatki so vzorno pripravljeni, saj z njihovim rokovanjem nismo imeli težav. Pred uporabo smo podatke dodatno obdelali. Najprej smo odstranili vse tiste povezave med kategorijami, pri katerih so bili zneski manjši od 1000 EUR. S tem smo želeli izpustiti podatke, ki so relativno majhni in, v kontekstu državnega proračuna, praktično nepomembni. Poleg tega smo na tak način zmanjšali velikost podatkovne zbirke in malenkost izboljšali odzivnost orodja, ki se ob velikem številu kategorij in tokov zmanjša.

Izračunali smo tudi vse skupne zneske po kategorijah. Nato smo pripravili uporabniku prijazen zapis zneskov, ki števila zaokroži in uporablja okrajšave za milijon in milijardo. Končno smo podatke preoblikovali v obliko, ki jo zahteva knjižnica za izris Sankeyevih diagramov (več v tem v nadaljevanju). Tako pripravljene podatke smo shranili za uporabo v nadaljevanju (opisana obdelava podatkov se izvede samo enkrat – orodje nato deluje na že obdelanih podatkih).

3.3.2 Tehnična izvedba. Za implementacijo Sankeyevih diagramov smo uporabili Pythonovo knižnico Plotly [6], ki ponuja številne interaktivne grafične prikaze in delo z njimi precej poenostavi. Plotly zahteva podatke o kategorijah in tokovih med njimi in iz njih avtomatično zgradi Sankeyev diagram.

Spletno aplikacijo smo zgradili z ogrodjem Dash [1] in jo objavili preko platforme Heroku [2]. Trenutno je na naslovu http://proracun.herokuapp.com na voljo verzija 0.3.

3.3.3 Oblikovalske odločitve. Ob oblikovanju diagramov smo morali sprejeti nekaj odločitev, ki so vplivale na uporabo in izgled diagramov. V prvi vrsti smo se odločali za funkcionalnost interakcij (glej razdelek 3.2). Pri spremembi pogleda se tako v primeru posebnega dela proračuna pokažejo tudi nadrejene kategorije, ker to nudi več konteksta, ki v splošnem delu proračuna ni tako pomemben.

Ime kategorije se pokaže, če je znesek kategorije vsaj 5 % vsote vseh kategorij v istem stolpcu. Podobno prikazujemo le prvih 30 znakov imena, celotno ime pa le ob interakciji. Obe meji (5 % in 30 znakov) smo določili empirično.

Vse kategorije so obarvane enako (svetlo sivo), razen bilanc, ki so temnejše, da bolj izstopajo. Tokovi so različnih barv, ki so določene tako, da so kategorije z istimi imeni vedno enako obarvane. To olajša razumevanje in primerjavo med različnimi leti proračuna. Z napisi na dnu prikaza, ki pojasnjujejo klasifikacijo, smo vnesli kontekst, ki pomaga pri orientaciji med spreminjanjem pogledov.

3.4 Razprava

Po začetnem testiranju uporabe, ki pa še ne vključuje prave uporabniške študije, lahko rečemo, da je Sankeyev diagram dober način za raziskovanje proračuna. Eden glavnih uvidov pri uporabi orodja je bil, da je servisiranje javnega dolga večja postavka od pričakovane (ker je poleg bilance odhodkov vsebovana tudi v računu financiranja, na preostalih vizualizacijah ne nastopa tako izstopajoče).

Interakcija omogoča "sprehajanje" po diagramu na različnih ravneh podrobnosti in v uporabniku zbudi željo po dodatnih informacijah, ki trenutno v vizualizacijo niso zajete. Te so na voljo le v obrazložitvah proračuna v datotečnem formatu PDF, kar otežuje njihovo morebitno dodajanje v aplikacijo.

Verjetno največja prednost takšnega prikaza je primerjava med posameznimi kategorijami in tokovi, ki je precej bolj intuitivna od obstoječih vizualizacij proračuna. Slabost je odzivnost, za katero bi si želeli, da bi bila boljša. Žal je to lastnost, ki se je ne da dovolj dobro predvideti in se izkaže šele v zadnjih fazah implementacije takšne aplikacije.

Uporaba knjižnice Plotly je zelo olajšala delo in zmanjšala čas, potreben za razvoj takšne aplikcaije. Vendar ima ta enostavnost uporabe za posledico (pre)malo nadzora nad končnim izgledom, ki bi ga želeli dodatno prilagoditi, a to ni mogoče. Moteča so tudi občasna prekrivanja imen v kategorijah (glej spodaj desno na sliki 1), ki se jim je pri interaktivnih vizualizacijah težko izogniti.

4 ZAKLJUČKI

Predstavili smo novo vizualizacijo proračuna Republike Slovenije s Sankeyevim diagramom, ki podpira interaktivnost in tako omogoča poglobljeno raziskovanje kategorij in denarnih tokov proračuna. Na ta način vizualiziramo tako splošni kot posebni del proračuna.

V nadaljevanju bi želeli poskusiti na isti način vizualizirati tudi razlike med dvema proračunoma. Tako bi lahko primerjali proračune dveh različnih let ali pa osnovni proračun z njegovim rebalansom.

ZAHVALA

Delo je nastalo v okviru raziskovalnega programa št. P2-0209, ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

LITERATURA

[1] Dash. 2020. Dash user guide. Dostopano 1. 9. 2020. https://dash.plotly.com.

[2] Heroku. 2020. Heroku. Dostopano 1. 9. 2020. https://www.heroku.com/home.

[3] Alex B. W. Kennedy in H. Riall Sankey. 1898. The thermal efficiency of steam engines. *Minutes of the Proceedings of the Institution of Civil Engineers*, 134, 278–312.

[4] Andy Kirk. 2016. *Data Visualization: A Handbook for Data Driven Design.* SAGE.

[5] Tamara Munzner. 2015. *Visualization Analysis and Design. AK Peters Visualization Series.* CRC Press.

[6] Plotly. 2020. Plotly Python open source graphing library. Dostopano 1. 9. 2020. https://plotly.com/python/.

[7] Edward R. Tufte. 2001. *The Visual Display of Qunatitative Information.* Graphics Press.

[8] Ministrstvo za finance Republike Slovenije. 2020. Državni proračun. Dostopano 1. 9. 2020. https://www.gov.si/podrocja/finance-in-davki/drzavni-proracun/.

[9] Ministrstvo za finance Republike Slovenije. 2020. Državni proračun 2020, Infografika. Dostopano 1. 9. 2020. https://www.gov.si/assets/ministrstva/MF/Proracun-direktorat/Drzavni-proracun/Sprejeti-proracun/Sprejeti-2020/Infografika_PRORACUN_2020.pdf.

[10] Ministrstvo za finance Republike Slovenije. 2020. Fiskalna in javnofinančna politika. Dostopano 1. 9. 2020. https://www.gov.si/teme/fiskalna-in-javnofinancna-politika/.

[11] Ministrstvo za finance Republike Slovenije. 2020. Proračun Republike Slovenije. Dostopano 1. 9. 2020. https://podatki.gov.si/dataset/proracun-republike-slovenije.

[12] Ministrstvo za finance Republike Slovenije. 2020. Proračun Republike Slovenije, Aplikacija APPrA. Dostopano 1. 9. 2020. https://proracun.gov.si/.

[13] Ministrstvo za finance Republike Slovenije. 2020. Vlada potrdila predlog rebalansa letošnjega državnega proračuna. Dostopano 1. 9. 2020. https://www.gov.si/novice/2020-08-30-vlada-potrdila-predlog-rebalansa-letosnjega-drzavnega-proracuna/.

MightyFields Voice: Voice-based Mobile Application Interaction

Jernej Zupančič
Jožef Stefan Institute
Ljubljana, Slovenia
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
jernej.zupancic@ijs.si

Miha Štravs
Faculty of Mathematics and Physics
Ljubljana, Slovenia
Faculty of Computer and
Information Science
Ljubljana, Slovenia
miha.stravs996@gmail.com

Miha Mlakar
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
miha.mlakar@ijs.si

ABSTRACT

We present MightyFileds Voice (MFVoice), a service and an extension of the MightyFields application that enables voice interaction with a mobile application. The user can issue voice commands for transitioning between application views and filling out the forms. Google speech-to-text engine is used to obtain text, which is then fed into the developed MFVoice service together with the structured application view representation. MFVoice service then returns appropriate action to take, which is executed by the Mighty Fields application extension. The MFVoice natural language understanding service was tested in real-life use cases, achieving 93% intent recognition accuracy, 88% entity recognition success when the system was used as intended. When no training to the user was provided, intent and entity recognition achieved 68% and 52% accuracy, respectively. Note that in case of no training provided, the users assumed general knowledge of the language semantics, which is out-of-scope for the current state-of-the-art research in natural language.

KEYWORDS

voice assistant, voice interaction, natural language understanding

1 INTRODUCTION

Interaction with devices by voice has become quite common in recent times. More known examples of applications allowing voice commands are voice assistants like Cortana [4] and Siri [1]. Voice interaction is attractive to users as it offers a hands-free application interaction and is therefore a desired feature in many applications. This feature is useful for people with spelling difficulties. It can also help those with physical disabilities who often find typing difficult. The proposed service is not used for two-way conversation, as in platforms such as the one from Rasa [3]. However, the part of the service used to recognize user's command, is very similar to the ones from other virtual assistants. The modifications applied take into account the specifics of the task at hand.

In this paper we focus on the task of filling out custom forms through the voice interaction. Here, a custom form is a small information gathering application, made for specific purpose, e.g., electric grid inspection form, or police report regarding an incident. Since the domain is open ended, i.e., each individual can make his or her own custom forms, the voice understanding feature cannot be specialized and has to work satisfactory in general setting. Three steps are performed to enable voice interaction. First, speech is transformed into text by using Google speech-to-text (STT) engine [2]. Second, approach from [5] is utilized to extract intent keywords. The full intended command is then inferred based on what the user is currently seeing on the screen and from the rest of the spoken words. Third, the recognized action is performed within the application itself.

In Section 2, an architecture of our service is presented. In Section 3, we present our MFVoice natural language understanding (NLU) service and show its implementation. We then explain the tests conducted on the service and their results in Section 4 and discuss them in section 5. We conclude the paper with a summary in Section 6.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information society '20, October 5–9, 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

2 MFVOICE ARCHITECTURE

MFVoice comprises several parts (Figure 1) that enable voice interaction:

(1) MF application itself: this is the main MightyFields application.
(2) MF agent: the program that enables programmatic access to the application view - reading and interacting.
(3) STT: a service that transforms spoken commands into text.
(4) MFVoice NLU service: the service that parses free text and returns structured information about recognized intent and entities.

3 THE MFVOICE NLU SERVICE

The MFVoice NLU comprises the following steps (Figure 2):

(1) Application view context processing
(2) Intent recognition
(3) Entity recognition

When the application context and transcription of the voice command are provided to the NLU application programming interface (API), the service first identifies possible actions to take, given the context, then it processes the context content, which in turn enables recognition of the intent and, finally, the entities. The so-obtained structured action data is then forwarded back to the MF agent, which can execute appropriate actions. In this section we will describe each of the MFVoice NLU parts in more details.

3.1 Application View Context Processing

The application view context provides structured data on the elements that are visible on the screen. This includes field labels, field IDs, possible values of fields (where applicable), interaction options, and available tabs for multi-page forms. Upon the API

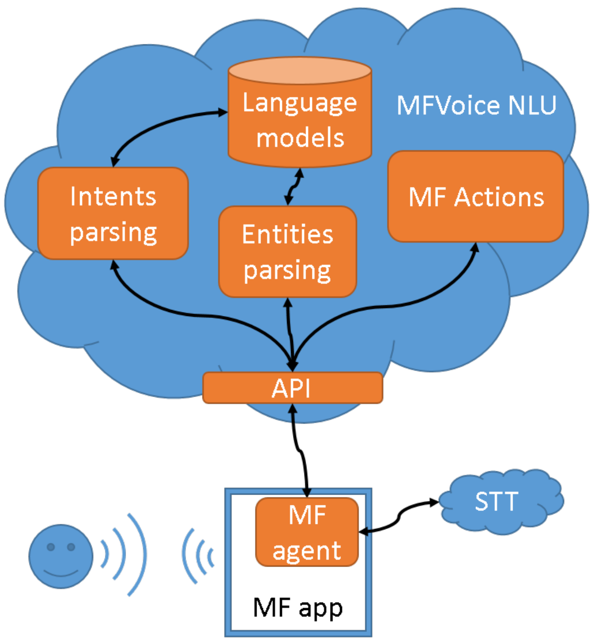


Figure 1: The MFVoice architecture overview

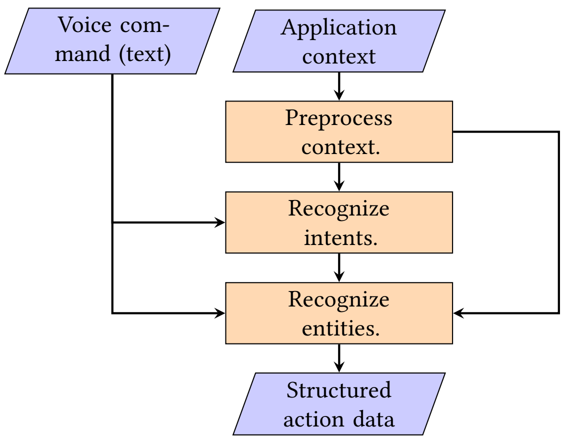


Figure 2: The MFVoice service processing pipeline

call, the application context is pre-processed. The text visible to the user is normalized and transformed into search friendly form that is used in intent and entity recognition.

The transformation is cached to speed-up its future use in subsequent steps. Least recently used cache is used, since user interacts with one application view as long as he or she does not fill-out the form in short time period.

3.2 Intent Recognition

Intent defines what the user wants the application to do. In the case of form filling, the following intents were identified:

(1) "Choose" an element from a list of elements. This is often used to pick the element from a dropdown or checklist elements.
(2) "Write" some text into a field. This is used for input or text-area elements.
(3) "Clear" the value of a field to delete a wrong value entry for any kind of element.
(4) "Tap" the field or graphical element. This is used to interact with buttons and navigate the application.

Due to non-existent training data, keyword-based intent recognition was utilized. For each intent a set of intent "key-phrases" were defined:

(1) "Choose": choose, pick, select, is
(2) "Write": is, write, input
(3) "Clear": clear, delete, remove
(4) "Tap": choose, go to, tab, click, pick

In the intent inference step, the score for each of the "key-phrases" is computed. If the highest score exceeds the predefined threshold, the corresponding intent is chosen [5]. To simplify the NLU pipeline only one intent and one field per utterance is allowed. This is especially problematic in voice assistants, since the users naturally communicate differently when talking than when writing. To resolve the disambiguation of one "key-phrase" being associated with multiple intents, the following order of intents is taken into account: "Choose", "Write", "Clear", "Tap". Further, since not every intent is possible with every field, the intent list is first filtered and only the intents that make sense in the current context are kept and iterated over. For instance, if the context comprises only of the input fields and navigation tabs, "Choose" does not make sense.

The "key-phrases" used for inferring the intent are tagged with the intent tag (IT), which is later used in the entity recognition step.

3.3 Entity Recognition

There are two types of entities present in our use-cases: "label entities" (e.g., name in *"The **name** is John."*) and "value entities" (e.g., John in *"The name is **John**."*).

Label entities are the labels of the fields in the user generated form. Since they are user generated, their value is not restricted. Their semantic meaning is sometimes harder to grasp automatically, since not much additional info is usually provided (for instance description). In general, the recognition of label entities cannot be learning-based, since the users are not expected to provide examples.

Value entities can further be divided into known value entities (button labels, the items in the drop-down lists or checkbox lists and similar) or unknown value entities (text input fields). The reasoning for known value entities is the same as for the label entities – they are user generated and cannot be learned in general. For the unknown value entities, the value corresponding to a certain label should be recognized from the free text, generated from the STT service. The unknown value entities can comprise one or several words.

Score, related to the probability of the entity appearing in the text, is used to recognize the label entities and known value entities, while a heuristic is used to recognize the unknown value entities.

3.3.1 Label and Known Value Entities Recognition. These entities are the ones set by the user in the form creation phase. To recognize the entities from free text, text similarity scores are applied and evaluated for each possible label or known value entity [5]. Only the entities with scores that exceed threshold are recognized and can be used in the next pipeline steps.

In some instances, the MFVoice NLU has to modify the text received by the STT engine. Common examples for this are:

(1) Letter-by-letter dictation transformation, e.g. in transcribed command "the form ID is 1 2 3 4 5 a" the empty spaces in the form ID have to be deleted so "12345a" is obtained.
(2) Zero padded numbers transformation, e.g. in transcribed command "the house number is 23", sometimes the drop-down values include only known value entity "023". Therefore, the preceding zeros have to be dropped when computing the text similarity.
(3) Numbers in text transformation, e.g. in transcribed command "pick option three", the number "3" is transcribed as "three". In those cases, the textual representation has to be transformed into a number.

The words that correspond to the highly-scored entities are tagged with the label entity (LE) or known value entity tag (KVE), to be later used in the unknown value entity recognition.

Some examples of label entity and known value entity recognition are:

(1)

form number	123456
LE	KVE

(2)

female
KVE

Note that when the user speaks command "female", the MFVoice NLU service recognizes the known value entity that belongs to the field with label "sex". Additionally, even without specifying the intent keyword, the application logic infers that the user wants to pick the option "female" in the field "sex".

Since the text similarity metrics are used for scoring the labeled and known value entities, in some cases the entities are not recognized correctly:

(1) Multi-word synonyms are not recognized, e.g. "city" ~ "place of living"

clear	the place of living
IT	OTHR

While the systems supports synonyms, they have to be manually entered by the form creator and are therefore less practical.
(2) Multiple occurrences of the same or very similar label entities or known value entities cannot be properly disambiguated. Consider, for example, a form that comprises house number field with possible value "4", and household size field also with possible value "4". User usually fills the form in a linear way, top to bottom. When the user encounters the first of the mentioned field, he or she may voice command "four". In this case, NLU service will provide two possible actions: "house number is 4" and "household size is 4".

3.3.2 Unknown Value Entity Recognition. The unknown value entity recognition is computed only when the intent "Write" is considered, since this is the only type of the"open" form field that allows for unknown values. The following heuristic is used to tag the unknown value entity (UVE):

(1) If IT tag is not present in the text, every word not tagged with LE is tagged with UVE.

age	31
LE	UVE

(2) If IT tag is present in the text, then begin tagging word to the left or to the right of LE-tagged word with OTHR tag. Stop if text-end or IT-tagged word are reached. Check if there is any remaining word:
(a) If there are remaining words, tag those with UVE tag.

his	name	really	is	John Doe
OTHR	LE	OTHR	IT	UVE

(b) If there are no remaining words, re-tag all the words to the right of LE tag with UVE tags.

insert	the	name	John Doe
IT	OTHR	LE	UVE

The previous steps capture the majority of the unknown value entity recognition cases. However, there are still commands that would not be understood by the MFVoice NLU service:

(1)

John Doe the	name	is	≠	John Doe	the	name	is
OTHR	LE	IT		UVE	OTHR	LE	IT

(2)

	John Doe really	is	his	name
	UVE	IT	OTHR	LE

≠	John Doe	really	is	his	name
	UVE	OTHR	IT	OTHR	LE

4 TESTING

The MFVoice NLU service was tested in two ways: laboratory testing and real-world testing. For laboratory testing, the text was entered into the service directly, bypassing the STT service. This way, the STT performance issues were ignored and only the recognition capability of the MFVoice NLU service was tested. The examples, however, were still obtained from the final MFVoice users. The test user was presented with an application screen and told to fill the form using only his or hers voice.

For the real-world testing, the users were given written instructions on how to use the app, however, no instruction on how to actually voice commands were given. First, the form was filled out using screen and keyboard interactions. Second, the field that a user wants for fill with a voice command was marked. Third, voice interaction was activated and the command was spoken. Fourth, the transcribed voice command, the context, and the marked item were stored for future analysis. We did not provide any examples on how to use MFVoice. This allowed us to research what the users actually expect from the system.

The forms used in testing included six free-text input field widgets (name, surname, age, settlement, street, house number), one radio widget with two options (gender: male, female), one checkbox field with five options (language: Slovene, Slovak, Spanish, Swedish, Sumerian), and four dropdown fields (country, settlement, street, and house number).

4.1 Laboratory Testing Set-up and Results

We have gathered 70 and 69 commands for application interaction in Slovenian and English languages, respectively. Laboratory testing is performed upon each git push to the code repository and is run within the continuous integration pipeline. This enables us to track the performance of the MFVoice NLU pipeline.

Table 1: Intent confusion matrix for commands in Slovenian

	write	choose	clear	tap	missing
write	22	0	0	0	2
choose	0	21	0	0	0
clear	0	0	2	0	0
tap	0	0	0	20	3

4.1.1 Intent Recognition. After each continuous integration pipeline run, the intent confusion matrix is computed. Table 1 is an example of the intent confusion matrix for voice commands in Slovenian for the last version of MFVoice. According to the matrix, the accuracy of intent recognition is above 90%. The only errors were the ones, where the system was not able to determine the item to be interacted with, which was labeled with the "missing" classification label.

4.1.2 Entity Recognition. For each command also the field labels and values recognized by the NLU service and the ground truth labels and values are compared. Examples where the NLU fails to recognize the label or value correctly are:

(1) "Age 26 years." Expected value: "26", got "26 years"
(2) "She is 26 years old." Expected label: "age", got nothing.
(3) "She lives in Ljubljana." Expected label: "Place", got nothing.

4.2 Real-life Testing Set-up and Results

We have gathered 172 spoken voice commands in the real-life setting in Slovenian. Unfortunately, there were only 86 commands that were labeled correctly by the test users and STT performed well there. STT issues occurred in 42 out of 172 cases (24%). These could either be result of too much background noise, command not being recorded properly, or just the problem with the STT service used for the Slovenian language. Incorrect user labeling occurred in 42 out of 172 cases (24%). The most common mistakes in those cases were: the user forgot to set the ground truth either by entering the value or choosing the item, the user obviously picked the wrong item (e.g., for command "the name is John" an item with label age was selected).

Out of 86 valid commands, 45 were recognized correctly and 41 incorrectly. For 23 cases the label value was completely missing and could not be inferred from the surrounding text (e.g., "John", "45", "Ljubljana"). For 18 cases the label value could be inferred from the surrounding text (e.g., "he is 23 years old", "she lives in Ljubljana"). In some cases (12) this would require some general reasoning about the words and their relations and in other the unknown value entity included additional text, e.g. "his name miki", was not recognized because of minor STT-engine mistakes (1), or the known value entity score was not high enough to be included (5). This results in 88% accuracy for entity recognition when the system was used as intended, 72% when the synonyms were assumed, and 52% when general knowledge of the language semantics was assumed.

Note that the testing was performed without some planned features implemented. The *Zero padded numbers transformation* and *Numbers in text transformation* steps were missing. The accuracy percentages should improve to 94%, 76%, and 56% for uses as intended, assuming synonyms, and assuming general knowledge of language semantics, respectively.

5 DISCUSSION

According to the results, the intent recognition process performs very good, despite the fact that it is only based on keyword recognition and the context processing. We do not think that any additional work would benefit the performance in this regard, with the exception of adding additional intent keywords, which will be obtained during the application usage.

After the user familiarizes with the way the MFVoice application works, also the named entity recognition performs well. Most of the errors were actually a result of a user expecting the system to be too advanced. All 43 incorrectly recognized entities were the result of MFVoice not being able to reason that, for instance, "John" is a person name. While this could be done for certain special cases, e.g. person names and geographic names, at the moment this cannot be solved in general. This is a result of letting the users to create their own forms, which are often very domain specific. In the future we will perform the testing of the system after users are given some basic training on how to use MFVoice. This should greatly improve the percentage of properly labeled instances and also help us uncover additional edge cases to be addressed by the entity recognition pipeline.

The MFVoice NLU was designed in a way to easily support multiple languages. In the current form, to support a new language, the translations of the intent keywords and language word vectors have to added. For certain languages the module for unknown value entity has to be adjusted, since the sentence syntax can be different. This enabled us to quickly add support for English, after Slovenian voice interaction performed well.

6 CONCLUSION

In this paper we presented our service that is used for filling forms with voice commands in a mobile application. While some operating system do include voice interaction, e.g., Cortana [4] and Siri [1]), their use in a dedicated application is limited. MFVoice enables more advanced voice interaction. MFVoice application first gets the text which was converted from speech by using the Google STT engine [2]. Then, the MFVoice NLU service uses keyword recognition and context preprocessing to infer the command the user intended. Because of the simplicity of the implementation, the service is less accurate when commands are voiced in the form of long and complex sentence. However, this simplicity does make the service more robust and accurate with commands voiced in concise form. We believe that users should have a comfortable user experience, after they get used to forming commands in a more concise manner.

ACKNOWLEDGMENTS

Comland d.o.o. funded the research presented in this paper.

REFERENCES

[1] Apple. 2020. Siri. https://www.apple.com/siri/. (2020).
[2] Google. 2020. Speech-to-text: automatic speech recognition. https://cloud.google.com/speech-to-text/. (2020).
[3] Rasa Technologies Inc. 2020. Rasa. https://rasa.com/. (2020).
[4] Microsoft. 2020. Cortana - your personal productivity assistant. https://www.microsoft.com/en-us/cortana/. (2020).
[5] Miha Štravs and Jernej Zupančič. 2019. Named entity recognition using gazetteer of hierarchical entities. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 768–776.

eBralec 4: hibridni sintetizator slovenskega govora

Jerneja Žganec Gros
Alpineon d.o.o.
Ulica Iga Grudna 15
1000 Ljubljana, Slovenija
jerneja.gros@alpineon.si

Miro Romih
Amebis d.o.o.
Bakovnik 3
1241 Kamnik, Slovenija
miro.romih@amebis.si

Tomaž Šef
Institut "Jožef Stefan"
Jamova cesta 39
1000 Ljubljana, Slovenija
tomaz.sef@ijs.si

POVZETEK

V članku predstavljamo nov sintetizator slovenskega govora eBralec4 (https://ebralec.si/). Razvit je bil povsem nov ženski glas »*Nadja eBralec*«, ki je razumljivejši in zveni bolj naravno od predhodnega ženskega glasu. Opisujemo zgradbo sintetizatorja govora, njegove module, jezikovne vire uporabljene pri razvoju ter potek izgradnje govorne zbirke za nov ženski glas.

KLJUČNE BESEDE

sinteza slovenskega govora, govorna zbirka, postopek sinteze slovenskega govora

1 Uvod

V članku predstavljamo nov sintetizator slovenskega govora eBralec4 (https://ebralec.si/). Pri razvoju smo izhajali iz obstoječe tehnologije, sintetizatorja govora za slovenski jezik *eBralec* [1], ki je bil razvit v okviru projekta Knjižnica slepih in slabovidnih, in je prvenstveno namenjen slepim in slabovidnim uporabnikom ter osebam z motnjami branja.

V okviru projekta CityVOICE smo identificirali več priložnosti za izboljšavo in nadgradnje *eBralca*, tako glede naravnosti kot tudi razumljivosti. V sodelovanju s skupino končnih uporabnikov smo pregledali in raziskali pomanjkljivosti obstoječega sintetizatorja govora ter zbrali dodatne želje končnih uporabnikov za izboljšave sintetizatorja govora, kar je rezultiralo v novem produktu *eBralec4*.

Kot poglavitna pomanjkljivost se je izkazala prvotna govorna zbirka, na kateri sloni delovanje izhodiščnega sintetizatorja govora. Ni zasnovana dovolj konsistentno (deloma neustrezni in spremenljivi snemalni pogoji) in robustno – v smislu zajema raznovrstnosti ciljnega besedišča, kar povzroča slabšo razumljivost sintetičnega govora ob črkovanju in izgovarjavi posebnih simbolov.

Identificirana je bila tudi spremenljiva kakovost sintetiziranega govora, ki izhaja iz sejne spremenljivosti ob snemanju izvorne govorne zbirke. Zaradi neustreznega dinamičnega obsega pri snemanju je manj uporabna za uporabo v mobilnih aplikacijah ob hrupnem akustičnem ozadju.

Zato smo velik del analize posvetili možnim izboljšavam pri gradnji nove govorne zbirke, ki omogoča boljše delovanje akustičnega modula. To še posebej velja za ženski glas, ki ga je zaradi fizikalne narave tudi sicer težje kvalitetno sintetizirati. Nova govorna zbirka za glas »*Nadja eBralec*« je bila posneta z branim govorom. To ustreza najpogostejšim oblikam rabe sintetizatorjev govora, lažje je izdelati transkripcijo, snemanje je bolj nadzorovano in predvidljivo. Pri spontanem govoru je namreč govorno zbirko težko fonetično in prozodično uravnotežiti.

Na osnovi analize delovanja izhodiščnega jezikovnega modula smo izboljšali pomensko analizo povedi in na novo razvili samodejno določanje vrste povedi, ki ima še posebej veliko težo tudi v postopku gradnje govorne baze.

Izpostavljene so bile tudi težave *eBralca* pri sintetiziranju kratkih besedilnih segmentov in posameznih simbolov, kar se kot najbolj moteče pokaže pri črkovanju, ki je bilo mestoma slabo nerazumljivo. Težavo smo rešili z uvedbo hibridnega pristopa k akustičnemu modeliranju govornega signala, kjer kratke segmente sintetiziramo z visoko razumljivo difonsko konkatenacijo govornih segmentov, daljše segmente pa z naravno zvenečimi parametričnimi reprezentacijami govornega signala s pomočjo prikritih Markovovih modelov.

Obstoječima glasovoma *eBralca*, moškemu glasu »*Renato eBralec*« in ženskemu glasu »*Maja eBralec*«, se je v novem produktu *eBralec4* pridružil novi in opazno bolj naravno zveneči ženski glas »*Nadja eBralec*«.

V članku opisujemo zgradbo sintetizatorja govora, njegove module, jezikovne vire, ki so bili uporabljeni pri njegovem razvoju, potek izgradnje nove govorne zbirke za ženski glas »*Nadja eBralec*« in postopek hibridnega akustičnega modela za generiranje govornega signala. Opisujemo tudi izboljšave pri jezikovni analizi vhodnega besedila.

2 Zgradba sintetizatorja

Naloga jedra sintetizatorja govora *eBralec* oziroma povezovalnega cevovoda je povezovanje sestavnih modulov sintetizatorja govora v enoten proces. Jedro sintetizatorja govora usklajuje delo posameznih delov sintetizatorja tako, da v ustreznem vrstnem redu vključuje oziroma kliče module sintetizatorja govora. Posamezni moduli pretvorbe zaradi pohitritve in večje paralelizacije procesov lahko hkrati delujejo v ločenih nitih.

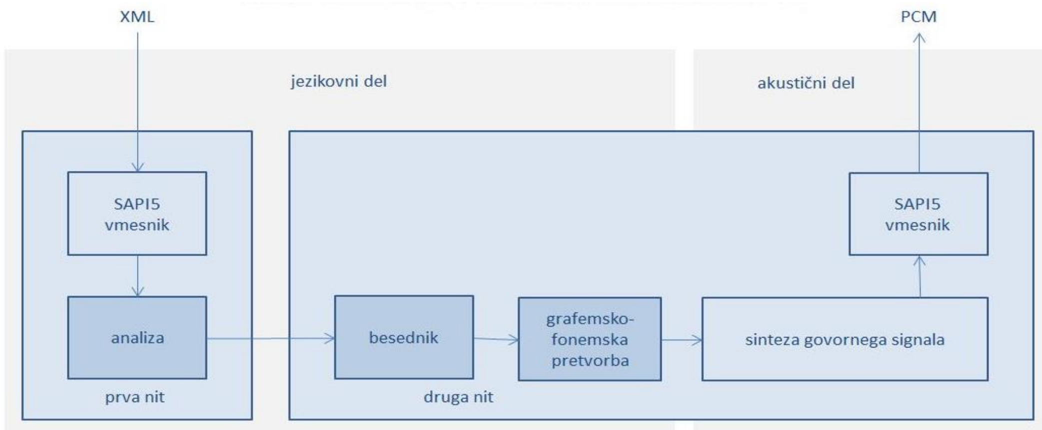
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

Zasnova jedra sintetizatorja govora *eBralec* je prikazana na sliki 1. Moduli, ki jih vključuje jedro *eBralca*, so: jezikovni analizator, besednik, modul za grafemsko-fonemsko pretvorbo in modul za sintezo govornega signala [1]. Na vhodu in izhodu se jedro sintetizatorja govora lahko poveže z ustreznim vmesnikom, npr. SAPI 5, s pomočjo katerega vhodno besedilo z morebitnimi dodatnimi ukazi spreminja v ustrezen govorni signal.

Vhodno besedilo sprva obdela jezikovni analizator, ki poskrbi za ustrezno predobdelavo vhodnega besedila ter razdvoumljanje izgovornih različic. Rezultat modula je zapis, v katerem so vsebovane vse potrebne informacije o izgovarjavi besed glede na njihovo pozicijo in pomen v vhodnem stavku oziroma povedi.

Modul besednik v odvisnosti od vhodnih nastavitev poskrbi za pretvorbo simbolov in števil v besede. Ti elementi so namreč zelo pogost sestavni del besedil, zato je njihovo pravilno izgovarjanje pomembno za razumljivost govora. Modul »grafemsko-fonemska pretvorba« poskrbi za pretvorbo v fonemski zapis.

Modul za »sintezo govornega signala« je zadolžen za oblikovanje prozodije in tvorjenje izhodnega govornega signala.



Slika 1: Shema jedra sintetizatorja govora.

3 Jezikovna analiza besedila

Jezikovna analiza uporablja podatke iz Amebisove jezikovne baze Ases [2]. Ta za slovenščino v tem trenutku vsebuje več kot 257.000 lem, ki vsebujejo 8,1 milijona oblik, od katerih je 5,7 milijona oblik dodatno opremljenih s podatki o izgovarjavi. Dodatno je za slovenščino v bazi še 36.000 zvez in 8.000 glagolskih predlog. Glagolske predloge podajajo informacije o vezljivosti glagola.

Jezikovni analizator mora narediti razrez besedila na povedi, stavke in besede, potem pa za vsako besedo določiti še ustrezno lemo in oblikoskladenjsko oznako. Ases ločuje leme, ki se različno izgovarjajo, npr. »téma« in »temà« predstavljata dve ločeni lemi. Jezikovni analizator deluje na podlagi pravil in podatkov iz jezikovne baze *Ases*, pri čemer so osnova glagolske predloge.

Na izboljšano delovanje sintetizatorja govora lahko jezikovno procesiranje vpliva predvsem s še boljšo analizo besedila, ki jo lahko uporabimo tako v postopku gradnje govorne baze kot tudi pri analizi besedila v fazi sintetiziranja govora

Na podlagi identificiranih težav delovanja jezikovnega modula obstoječega sintetizatorja smo veliko pozornosti posvetili možnim izboljšavam jezikovnega analizatorja. Posebej pomembni med njimi sta izboljšanje pomenske analize povedi in *določanje vrste povedi*, ki ima še posebej veliko težo tudi v postopku gradnje govorne baze, gl. poglavje 4.

Raziskali smo tudi možnost pohitritve odzivnosti oz. latence jezikovnega analizatorja s pomočjo postopkov za učinkovito računalniško predstavitev leksikalnih jezikovnih virov, ki jih razvijamo v okviru projekta OptiLEX. Pri tem rešujemo vrsto problemov, kot so: zahteva po delovanju v realnem času, zahteva po kompaktnem zapisu jezikovnih virov ter zahteva po majhnem odtisu zapisa jezikovnih virov v delovnem pomnilniku [4].

3.1 Samodejno določanje vrste povedi

Pri izbiri optimalnih fonetično in drugače uravnoteženih besedilnih predlog za govorno bazo smo se posvetili izboljšanemu označevanju povedi, predvsem označevanju in določanju vrste povedi, kjer smo poleg klasične metode s pomočjo pravil analizirali tudi možnost določanja vrste povedi s pomočjo različnih metod strojnega učenja. Razviti postopek smo uporabili tudi v izboljšani različici jezikovnega analizatorja.

Za potrebe določanja vrste povedi, predvsem večstavčnih, smo definirali ustrezne zapise. Osnovni zapis, prilagojen zdajšnjemu zapisu analize povedi, je kompleksen. Poved je zapisana v posebnem meta jeziku, ki vsebuje vse informacije, ki jih lahko izluščimo iz povedi na osnovi avtomatske stavčne analize.

Tak zapis omogoča združen zapis večstavčnih povedi, v katerem nastopajo tudi vse stavčne odvisnosti. Stavki v povedi namreč lahko nastopajo kot priredja, soredja ali podredja. V primeru podredja pa je navzoča tudi informacija glede njihove odvisnosti, torej ali gre za prilastkov, osebkov, predmetni, prislovni ali kateri drugi odvisnik.

Poleg tega daljšega zapisa smo definirali tudi skrajšan, poenostavljeni zapis, ki podaja informacijo o vrsti povedi, ki smo ga uporabili pri izbiri končne množice izbranih povedi.

Ker izboljšani stavčni analizator svojo analizo zapiše v daljšem zapisu, smo razvili pretvornik iz tega zapisa v poenostavljeni zapis, ki ohrani le najbolj pomembne podatke o tipu strukture povedi. Primer tega zapisa za poved "*Miha, ki je bil lačen, je pojedel malico*." je "*[[-gp|[[-r-|]]]]*".

S pomočjo pretvornika smo vhodni množici povedi dodali informacijo o *vrsti povedi*, ta pa je v postopku izbire povedi služila kot eden izmed parametrov pri izbiri in uravnoteženju ciljnega števila povedi. Analizator, ki zapiše analizo povedi z informacijo o vrsti povedi v daljši zapis, smo realizirali s pomočjo pravil in z metodami strojnega učenja. V analizator smo vgradili rešitev, ki je ob evalvaciji dala najboljše rezultate [5].

4 Govorna zbirka CITYVOICE

V osrednjem delu analize smo izhajali predvsem iz zaznanih pomanjkljivosti obstoječega sintetizatorja govora, *eBralca*, ki bi jih bilo mogoče izboljšati, ter zbranih uporabniških zahtev ter identificirali priložnosti za razvoj izboljšanega produkta. Sem vsekakor sodijo težave zaradi neoptimalne govorne baze, zato smo velik del analize posvetili možnim izboljšavam pri gradnji nove govorne baze, ki pomembno vpliva na izboljšano delovanje akustičnega modula sintetizatorja govora.

Izbira velikosti govorne zbirke je posledica kompromisa med želenim številom variacij glasov oz. njihovim pokritjem na eni strani ter časom in stroški, vezanimi na razvoj, na drugi strani. Upoštevali smo tudi čas za kasnejše preiskovanje govorne zbirke in potreben prostor za njeno hranjenje. Najpomembnejši preostali dejavniki, ki smo jih upoštevali pri snovanju nove govorne zbirke za sintezo govora, so: izbira vsebine posnetkov, izbira govorcev, postopek snemanja in označevanje posnetkov.

Izbor povedi za govorno bazo poteka na osnovi večjega števila kriterijev, med katerimi so pokritost osnovnih govornih enot, uravnoteženost dolžin, tipov in vrst povedi, pravilna fonetična transkripcija itd [6]. Med njimi bi posebej omenili *vrsto povedi*. Ta omogoča bolj natančno modeliranje prozodije, ki je pomembna za naravnost sintetičnega govora. Ena od glavnih težav starega postopka je bila v tem, da nismo imeli orodja za avtomatsko določanje vrste povedi, kar bi lahko bistveno pohitrilo in izboljšalo izbiro povedi.

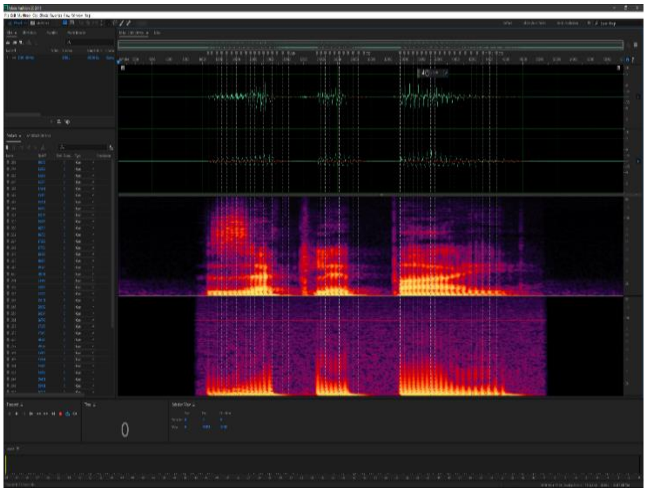
Pri izbiri optimalnih fonetično in drugače uravnoteženih vsebin (povedi) smo se zato v veliki meri posvetili izboljšanemu označevanju povedi, predvsem označevanju in določanju *vrste povedi*, kjer smo poleg klasične metode s pomočjo pravil analizirali tudi možnost določanja vrste povedi s pomočjo različnih metod strojnega učenja, kot je to opisano v poglavju 3.1.

Za razliko od avtomatskega določanja *tipa* povedi, ki je v večini primerov odvisna od končnega ločila (trdilni, vprašalni, velelni) je določanje *vrste* povedi precej bolj zapleteno. Če je poved enostavčna, težav ni. Če pa je poved večstavčna, je potrebna zahtevna analiza povedi in vseh njenih stavkov za določitev njihove odvisnosti. Ti lahko nastopajo kot priredja, soredja ali podredja.

Snemanje govornega gradiva je potekalo ob prisotnosti izkušenega snemalnega operaterja z namenom, da se je preprečilo neustrezne izgovarjave besedilnih predlog in napake pri snemanju govora, gl. sliko 2.

Slika 2: Govorka med snemanjem oz. branjem pripravljenega gradiva v tonskem studiu. V ozadju tonski tehnik, ki med snemanjem na zaslonu spremlja tako signal laringografa Lx kot tudi mikrofonski signal Sp.

Govorcu smo pred snemalnimi sejami podali ustrezna navodila in ga zaprosili, da povedi prebira razločno in enakomerno hitro. Med branjem besedila so imeli govorci nameščene elektrode laringografa, s katerimi smo spremljali nihanje njihovih glasilk zaradi lažjega kasnejšega označevanja osnovnih period govornega signala, gl. sliki 1 in 2. Uporabili smo tri nivoje anotacij oz. prepisov govorjenega besedila: grafemski prepis, fonetični prepis in prozodijske oznake (slika 3).



Slika 3: Primer govornega signala z označenimi osnovnimi periodami. Zgoraj je govorni signal posnet z mikrofonom, sledi signal laringografa Lx in spektralna prikaza obeh signalov. Navpične črte predstavljajo oznake period govornega signala.

Uporabili smo tri nivoje anotacij oz. prepisov govorjenega besedila: grafemski prepis, fonetični prepis in prozodijske oznake (slika 3).

5 Akustično modeliranje govora s hibridnim postopkom

Specifikacije končnih uporabnikov so narekovale hitro odzivnost sintetizatorja govora ter kompaktno velikost pomnilniškega prostora, potrebnega za namestitev ter delovanje sintetizatorja govora.

To je ponovno narekovalo izvedbeno različico končnega sintetizatorja govora, ki, podobno kot pri *eBralcu*, temelji na parametrični predstavitvi zakonitosti govora v slovenskem jeziku [1] z uporabo prikritih Markovovih modelov PMM [7,8]. Teh zakonitosti se sintetizator govora nauči samodejno na podlagi obsežnega učnega govornega korpusa, ki je bil posebej posnet v te namene, in ki vključuje relevantne akustične ter prozodijske fenomene, ki so značilni za govorjeno slovenščino.

Sinteza govora z uporabo prikritih modelov Markova (PMM) ima v primerjavi z bolj klasičnimi postopki tvorbe govora, pri katerih govor tvorimo z »lepljenjem« krajših ali daljših govornih izsekov, nekaj privlačnih prednost, saj za zadovoljivo kakovost govora potrebujemo razmeroma majhno govorno zbirko (zadošča že ura ali več posnetega govora). Nadalje omogoča enovito, kakovostno in sočasno modeliranje akustičnih in prozodičnih značilnosti govora. Omogoča tudi zgoščen zapis akustičnega in prozodijskega modela govora, saj za tvorbo govora ni treba hraniti celotne izvorne govorne zbirke.

Po drugi strani pa imajo sistemi PMM tudi nekatere slabosti. Govor je lahko na trenutke nekoliko manj razumljiv. Govor ima lahko ponekod značilen »robotski« prizvok, ki je posledica parametrizacije govornega signala.

Podrobna analiza uporabniške izkušnje slepih in slabovidnih uporabnikov *eBralca* je pokazala, da je še posebej slabo razumljiva sinteza govora *krajših* besednih enot, kot je denimo črkovanje, ki ga ta skupine končnih uporabnikov zelo pogosto uporablja. Slepi in slabovidni uporabniki namreč za uspešno uporabo računalnika uporabljajo t. i. bralnike zaslona, programe, ki s pomočjo sintetizatorja govora uporabnikom sporočajo informacije o tem, kaj se prikazuje na ekranu.

Za točno sliko ekrana zelo pogosto uporabijo *branje v načinu črkovanja*, ki besede bere črko po črko, oz. znak po znak. Pri tem je potrebna velika hitrost branja oz. izgovarjanja, pri čemer metoda PMM ni najbolj uporabna, ker je premalo odzivna ter rezultira v manj razumljivih kratkih izoliranih segmentih.

Ker je postopek sinteze s pomočjo PMM manj primeren za uspešno sintezo kratkih govornih segmentov, smo se odločili za razvoj unikatnega **hibridnega akustičnega modela**, ki omogoča kakovostno sintezo govora *kratkih* govornih segmentov s pomočjo difonskega sintetizatorja govora z uporabo konkatenacije osnovnih govornih segmentov z metodo TD-PSOLA [9,10], daljši govorni segmenti pa so generirani z uporabo pristopa s pomočjo prikritih Markovovih modelov PMM.

6 Zaključek

V prispevku smo predstavili zasnovo in izvedbo novega visokokakovostnega sintetizatorja govora za slovenski jezik, *eBralec4*. Za samodejno tvorjenje govora smo uporabili optimizacijo postopka pridobivanja govornih jezikovnih virov v kombinaciji z napredno parametrično predstavitvijo govora z modeliranjem govora s pomočjo prikritih Markovih modelov ter difonsko konkatenacijsko sintezo govora za sintezo krajših govornih segmentov, predvsem pri črkovanju.

Izdelali smo govorno zbirko za nov ženski glas, »*Nadja eBralec*«. Pri izdelavi govorne zbirke smo posebno pozornost namenili določanju optimalnih pogojev za snemanje ter določanju optimalnih fonetično in drugače uravnoteženih besedilnih vsebin, pri čemer smo dodali raznovrstnost povedi glede na novo razviti postopek samodejnega določanja zvrsti povedi.

Zahvala

Razvojno raziskovalno delo je bilo delno financirano v okviru projekta CityVOICE s strani Republike Slovenije in Evropske unije iz Evropskega sklada za regionalni razvoj, in sicer v okviru »Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014-2020«. Raziskave učinkovitega zapisa jezikovnih virov je delno sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije v sklopu aplikativnega raziskovalnega projekta OptiLEX (L7-9406).

LITERATURA IN VIRI

[1] Jerneja Žganec Gros, Boštjan Vesnicer, Simon Rozman, Peter Holozan, Tomaž Šef, 2016. Sintetizator govora za slovenščino eBralec. Konferenca Jezikovne tehnologije in digitalna humanistika, Ljubljana.

[2] Špela Arhar in Peter Holozan. 2009. ASES – leksikalna podatkovna zbirka za razvoj slovenskih jezikovnih tehnologij. V Mikolič (ur.). Jezikovni korpusi v medkulturni komunikaciji. Koper: Založba Annales.

[3] Peter Holozan. 2004. Uporaba glagolskih predlog pri strojnem prevajanju. V: Zborniku Konference JEZIKOVNE TEHNOLOGIJE 2004, uredila T. Erjavec in J. Žganec Gros, str. 128. Ljubljana.

[4] Žiga Golob, Jerneja Žganec Gros, Mario Žganec, Boštjan Vesnicer, "*FST-based pronunciation lexicon compression for speech engines*" International Journal of Advanced Robotic Systems, zv. 9, 2011..

[5] P. Holozan, M. Romih, S. Rozman, 2019, R2.2 – Zasnova govorne zbirke, projektno poročilo CityVoice – govorne tehnologije z naprednimi jezikovnimi viri.

[6] Tomaž Šef, Miro Romih, Jerneja Žganec Gros, 2019. Izdelava govorne zbirke za sintezo slovenskega govora, Informacijska družba IS 2019.

[7] T. Toda in K. Tokuda. 2007. A speech parameter generation algorithm considering global variance for HMM-based speech synthesis. IEICE Transactions Inf. Syst.

[8] H. Zen in H. Sak. 2015. Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis. V: Proceedings of the ICASSP, str. 4470–4474.

[9] Jerneja Žganec Gros in Mario Žganec. 2008. *An efficient unit-selection method for concatenative text-to-speech synthesis systems*. CIT, zvezek. 16, št. 1, str. 69-78.

[10] C. Hamon, E. Moulines, F. Charpentier, (1989), “A Diphone System Based on Time-Domain Prosodic Modifications of Speech”, Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP 89, S5.7, str. 238–241.

Sound 2121: The Future of Music

Jordan Aiko Deja
jordan.deja@famnit.upr.si
University of Primorska
UP FAMNIT
Koper, Slovenia

Nuwan Attygale
nuwan.attygalle@upr.si
University of Primorska
UP FAMNIT
Koper, Slovenia

Klen Čopič Pucihar
klen.copic@famnit.upr.si
University of Primorska
UP FAMNIT
Koper, Slovenia
Fakulteta za Informacijske Študije
Novo mesto, Slovenija

Matjaž Kljun
matjaz.kljun@famnit.upr.si
University of Primorska
UP FAMNIT
Koper, Slovenia
Fakulteta za Informacijske Študije
Novo mesto, Slovenija

Figure 1: Concept: We see a future where we no longer need tangible interfaces. Rather humans would let go of these interfaces to give way to a more seamless music interface.

ABSTRACT

Music has always been an integral part of our society since the prehistoric times. For the past five centuries, music instruments have been perfected and the industry is nowadays worth billions of dollars. With recent innovations in computer interfaces, music information retrieval and artificial intelligence, playing music is not in the sole domain of humans anymore. Thus we are faced with the questions: "What really is music? What is the future of music? How will we consume music a hundred years from now?" In this paper, we shortly present how music has been consumed throughout history and how we imagine it a century from now. We make a wild speculation about the future of music and its interface, while encouraging the discussion regarding these visions.

KEYWORDS

music, interface, interaction design, sound, future

1 INTRODUCTION

Music is considered to be culturally universal [2, 17] and present across all parts of the globe, reshaping the ways human live, express themselves and convey emotions [9, 13]. Humans have been expressing themselves through music

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

for a very long time. It is believed that music originated from naturally occurring sounds and rhythms that humans echoed by merging them in patterns, making repetitions while changing tonality using their voice, hands clapping [11], and smacking stones, sticks and other objects around them [12]. For example, one of such ambient sound is rain, that has a calming and a relaxing effect, even now since early humans felt safe during the rain while predators do not hunt [1]. Music has also helped humans in terms of survival, forging a sense of group identity and mutual trust [4].

The voice box, which allowed humans to sing, first emerged about a million years [13] ago and they learned how to use it around 530,000 years ago [14]. The voice box is considered the first music instrument. Besides voice and hands, the earlier instruments were the objects found in the environment such as sticks and stones. Some authors argue that since the oldest instruments found are so sophisticated (such as over 40 thousand year-old bone flutes [4, 8, 19]) there must have been less sophisticated instruments used by humans before [4, 14]. Nevertheless, the instruments the humans made and used have rapidly evolved together with the complexity of music compositions in the last couple of centuries. In this period a variety of string, brass, percussion and woodwind instruments have evolved from earlier less sophisticated ones [3].

As newer technologies are introduced, other ways of creating, producing, interacting with and even sharing music [18] are also taking place. MIDI interfaces, electric guitars and synthesizers are just some of the devices made of circuits that imitate traditional music instruments, and can be connected to the computer. Novel algorithms, music information retrieval (MIR) [5, 6] and artificial intelligence (AI) techniques allow us to work with and create new music content. With the advent of social platforms, sharing music on a grand scale has become a norm.

Throughout this evolution one of the main components of music is expressing and generating emotions. Changes in vocal parameters occurring during speech as well as singing have been shown to effect our state of emotions [10]. It has been also confirmed that sadness, happiness and other emotions can be communicated to listeners by music composers. As such, music is considered as a popular and easily-applicable means for triggering emotions [10] and is globally consumed by everyone. We listen to music in order to make us happy, sad, to reminisce or reflect on our emotions.

This paper attempts to share the authors' visions on how humans will consume and interact with music in the future. We present our position based on the trends in how music instruments and music consumption have evolved throughout the history. These visions have also emerged from shared ideas in our small crowd-sourcing study we conducted online. We present two scenarios of how future music consumption might look like. Lastly, we present questions and challenges that provoke discussions involving usability, security, intellectual property and many other relevant key topics on music.

2 RE-IMAGINING MUSIC AND THE MUSIC INTERFACE

Humans create and consume music for four different purposes: (1) dancing as a social exercise, (2) providing a common form of personal or community entertainment, (3) communicating ideas and emotions and (4) having and celebrating rituals and other activities [13]. While these purposes come in handy for a variety of music activities, this position paper is focusing on music listening only. This is present in (1) where listening is a shared experience, as well as in (2) and (4) where listening can be a shared and also a personal experience.

When looking at how music listening has evolved in history, we can envision human tribes gathered around the fire where one or several members performed a music piece. This type of music listening has been present for a long period of time and even nowadays people gather in live concerts to listen to music. With recordings, music has moved to people's homes and the group has been reduced to family members and friends and listening become more personal. The headphones enabled users to experience music individually and the Walkman enabled us to do it on-the-go. Smartphones and internet have expanded the instant availability of music but the consumption remained mainly personal. The advances in virtually reality (VR) and augmented reality (AR) have made personal music listening an immersive experience with lights and visualisations augmenting music. Looking at how listening to music has moved closer and closer to us with in-ear headphones that we even try to insert as close as possible to our body, it is not far-fetched if our vision is that listening to music will move inside our heads.

Music is not just about sounds as it is also about rhythmic vibrations; for example, it has been noted that the part of the brain responsible for hearing, works perfectly in deaf people as well [16]. In order to feel music, we do not need to hear it but rather receive the vibrations to the hearing region of the brain. Because of this, we envision a future where we do not need external devices (such as headphones or speakers) to be able to hear music, enjoy concerts, etc. Rather we will be listening to music within our brain in a seamless way. Currently, researchers are already experimenting with micro-controllers plugged into the brain and we envision having similar devices plugged into the hearing part of our brain. Sound signals will be delivered straight into our auditory cortex. People will no longer have to depend on their ears to listen and hear things. As such, people could enjoy music even while spacewalking, diving, skiing or surfing.

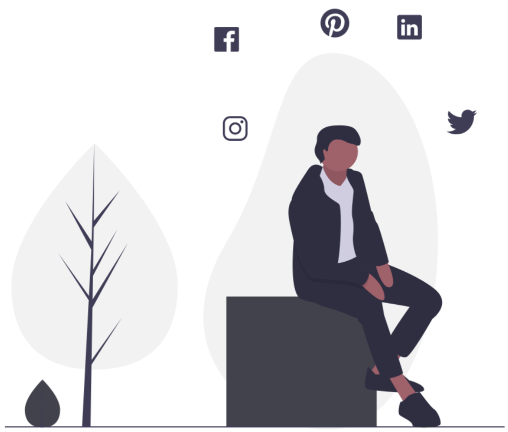


Figure 2: Concept: Humans do not need tablets, mobile devices, or digital walls. Ideas and concepts are conceived in and are played into our brains from the surrounding objects.

At the same time, we envision a future where biological and artificial objects around us will be connected to the cloud [15] where they will have access to a superb computing power. These objects will be equipped with for example nano-chips that will allow them to be part of the global link of information and capable of moving it depending on the needs. This is partly also a vision of IoT [20], which we are expanding to music listening. These technologies will allow humans and objects to telepathically communicate.

In the future humans will be able to amplify their emotions by the music naturally produced by the objects surrounding them (see Fig 2. Traditionally, there are two ways on how music becomes a gateway for our emotions. If we feel sad, we wish to hear music so we can reflect, dive deeper and understand the sadness that we feel [emotions going in]. This experience gives us lessons on how to manage our emotions, and how to grow stronger. At times, we may feel sad so we wish to hear music in order to improve our mood [emotions going out] and spend the better part of our days. In our envisioned interface, humans can create gateways for their emotions with music.

Algorithms will design and produce rhythms in on-the-fly and have them played via vibration by these nearby objects (moving on their own). Humans will simply need to think of their emotions and sounds, and the objects near them will seamlessly produce the vibrations recreating these sounds. Objects around us, will produce a unique rhythm, providing a new definition of audio augmented reality. Humans will get to enjoy their favorite sounds and rhythms through this seamless interface and played directly in their minds. We believe that through this interface, humans will be able to listen and consume music, various melodies and sounds while being able to perform their daily tasks at the same time.

As humans are connected to the global highway of information, unobtrusive sensors no longer need to detect and distinguish the current affect that they feel. Because of their neurological connection to the world, their emotions are easily read and "felt" by the objects around them. Similar to how empathetic spaces that are context and emotion aware, objects nearby will be act as local producers of music to either amplify or address the emotions that humans are feeling.

3 DESIGN SCENARIOS

To better explain how we re-imagined this music interface, we describe two design scenarios with our vision in different contexts.

Figure 3: Concept: Listening to music while surfing in the wide ocean will no longer require waterproof music gear. Rather, natural elements that are interlinked together create vibrations that humans can hear. Humans can finally achieve a non-obtrusive way of listening music while enjoying their wet hobbies.

Surfing. It is a lovely sunny and windy day. Cuauthli decides that these are perfect conditions to go surfing (as seen in Fig 3). While doing it he likes to feel the adrenaline rush with the sound of rock music. In 2020 Cuauthli would have to wear water-proof in-ear headphones tightly plugged into his ears to prevent them from falling off. This would prevent him to hear the surroundings. However, when surfing he also has to hear the surroundings for his safety. In order to do this, he would have to balance spatial awareness and enjoy at the same time, which takes a lot of effort [7]. In 2121 he will not have this problem anymore since, not only can he hear his preferred rock music, but the music blends with the sound of the environment around him. In addition, if Cuauthli wants to listen to the environment, the algorithm understands this and can mute the music just through his thoughts. This can be done using two approaches. First, Cuauthli can listen to his favorite track in full volume and when he wants to listen to the background noise, a chip inside his head can understand this and allows the environment sound to be heard. Second, the noise of the environment around him can be used as an input and then be used to create a new sound that blends with his taste and with the noise around him. These can be based on Cuauthli's favorite tunes and algorithms produce a specific tune that fits his current preferences and allows him to not lose the connection with the environment around him.

Amplifying emotions at the blink of an eye. It is a rainy day and Cuauthli is sitting by the window, thinking about his loved one. Since he is in Germany on a research visit, he misses her dearly. Cuauthli would love to get lost in his thinking about her. He then decides to listen to a song, which helps him reflect on his feelings for her and on his current mood. The music helps him bringing back the the memories. This is done by reducing other background noise as inputs and allows him to focus on the memories that are in his brain. After a while, the rain stops falling and Cuauthli needs to go back to work, but is feeling somewhat depressed. He thinks of a happy and exciting song, which starts playing and helps him to focus on his work as well as changes his mood. The algorithms and his neurological link take care of processing his thoughts and produce the sounds that he needs to hear.

4 CONCLUSION

The visions and scenarios we presented come with their respective issues and challenges in implementation and in policy design. If we imagine a natural and seamless interface, evaluating its usability will introduce a new paradigm for HCI researchers. Will existing models such as Fitts' Law (which has always worked on any interface developed - mechanical, digital, virtual) still work in neurological links managed by our seamless thoughts? The intangible interaction provided by this *"online network"* could potentially blur concepts such as piracy and intellectual property. As music is composed by ubiquitous algorithms connected to our personal thoughts and feelings, are all our emotions and the music that are generated by them considered unique and shareable? These, among many others, are interesting questions that we leave to our readers as we re-imagined a music interface. While these visions appear to be very far from reality, we are only left with our own thoughts to begin with and maybe hopefully in the not so near future too.

REFERENCES

[1] Mitsuko Aramaki, Richard Kronland-Martinet, and Sølvi Ystad. 2017. *Bridging People and Sound: 12th International Symposium, CMMR 2016, São Paulo, Brazil, July 5–8, 2016, Revised Selected Papers.* Vol. 10525. Springer.

[2] Patricia Shehan Campbell. 1997. Music, the universal language: Fact or fallacy? *International Journal of Music Education* 1 (1997), 32–39.

[3] Nicholas J Conard. 2009. A female figurine from the basal Aurignacian of Hohle Fels Cave in southwestern Germany. *Nature* 459, 7244 (2009), 248–252.

[4] Nicholas J Conard, Maria Malina, and Susanne C Münzel. 2009. New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 7256 (2009), 737–740.

[5] J Stephen Downie. 2003. Music information retrieval. *Annual review of information science and technology* 37, 1 (2003), 295–340.

[6] J Stephen Downie. 2008. The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology* 29, 4 (2008), 247–255.

[7] Dominik Fuchs. 2018. Dancing with Gravity—Why the Sense of Balance Is (the) Fundamental. *Behavioral Sciences* 8, 1 (2018), 7.

[8] Thomas Higham, Laura Basell, Roger Jacobi, Rachel Wood, Christopher Bronk Ramsey, and Nicholas J Conard. 2012. Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: The radiocarbon chronology of Geißenklösterle. *Journal of human evolution* 62, 6 (2012), 664–676.

[9] Patrik N Juslin and John A Sloboda. 2001. *Music and emotion: Theory and research.* Oxford University Press.

[10] Mattes B Kappert, Alexandra Wuttke-Linnemann, Wolff Schlotz, and Urs M Nater. 2019. The Aim Justifies the Means—Differences Among Musical and Nonmusical Means of Relaxation or Activation Induction in Daily Life. *Frontiers in Human Neuroscience* 13 (2019), 36.

[11] Jamie C Kassler. 1987. The dancing chimpanzee: A study of the origin of music in relation to the vocalising and rhythmic action of apes. *Musicology Australia* 10, 1 (1987), 79–81.

[12] Jeremy Montagu. 2014. *Horns and Trumpets of the World: An Illustrated Guide.* Rowman & Littlefield.

[13] Jeremy Montagu. 2017. How music and instruments began: a brief overview of the origin and entire development of music, from its earliest stages. *Frontiers in Sociology* 2 (2017), 8.

[14] Iain Morley. 2013. *The prehistory of music: human evolution, archaeology, and the origins of musicality.* Oxford University Press.

[15] Elon Musk et al. 2019. An integrated brain-machine interface platform with thousands of channels. *Journal of medical Internet research* 21, 10 (2019), e16194.

[16] ABC Online. 2001. Music - good vibrations for deaf. http://www.abc.net.au/science/articles/2001/11/28/426276.htm. Accessed: 2020-09-10.

[17] Charles Seeger. 1971. Reflections upon a given topic: Music in universal perspective. *Ethnomusicology* 15, 3 (1971), 385–398.

[18] Amy Voida, Rebecca E Grinter, Nicolas Ducheneaut, W Keith Edwards, and Mark W Newman. 2005. Listening in: practices surrounding iTunes music sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 191–200.

[19] Nils Lennart Wallin, Björn Merker, and Steven Brown. 2001. *The origins of music.* MIT press.

[20] Robert Weisman. 2004. The Internet of things: Start-ups jump into next big thing: tiny networked chips. *The Boston Globe* (2004).

Ohranjanje kulturne dediščine s pomočjo navidezne in obogatene resničnosti

Cultural heritage preservation through virtual and augmented reality

Marko Plankelj
Univerza v Mariboru,
Fakulteta za elektrotehniko,
računalništvo in informatiko
Maribor, Slovenija
marko.plankelj@student.um.si

Niko Lukač
Univerza v Mariboru,
Fakulteta za elektrotehniko,
računalništvo in informatiko
Maribor, Slovenija
niko.lukac@um.si

Selma Rizvić
Univerza v Sarajevu,
Fakulteta za elektrotehniko
Sarajevo, Zmaja od Bosne, bb.,
Bosna in Hercegovina
selma.rizvic@etf.unsa.ba

Simon Kolmanič
Univerza v Mariboru,
Fakulteta za elektrotehniko,
računalništvo in informatiko
Maribor, Slovenija
simon.kolmanic@um.si

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

© 2020 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.23

Povzetek

Kulturna dediščina izginja zaradi različnih dejavnikov in pri njenem ohranjanju si v zadnjih letih vse pogosteje pomagamo s sodobnimi tehnologijami, ki omogočajo njeno digitalizacijo. Kot primer dobre prakse predstavimo uporabo navidezne in obogatene resničnosti. Vedno pogosteje se uporablja tudi mešana resničnost, ki združuje virtualne objekte in resnično okolje. V članku predstavljamo možnost virtualne razstave muzejskih eksponatov s pomočjo očal Microsoft HoloLens in igralnega pogona Unity. V ta namen smo v okviru diplomske naloge ustvarili aplikacijo, ki omogoča interakcijo s šestimi artefakti iz časov Rimljanov, najdenih na štirih različnih nahajališčih na Balkanu. V članku predstavljamo prednosti in slabosti, ki jih taka predstavitev nudi uporabniku in kako lahko le-to uporabimo za ohranjanje kulturne dediščine.

Ključne besede

Kulturna dediščina, mešana resničnost, obogatena resničnost, navidezna resničnost, Microsoft HoloLens

Abstract

Cultural heritage is disappearing due to various factors. In recent years, we have increasingly used modern technologies together with its digitalization in order to preserve it. As an example of good practice, we present the use of virtual and augmented reality. Increasingly, mixed reality is also used combining virtual objects and a real environment. In this article, we present the possibility of virtual exhibition of museum artefacts with the help of Microsoft HoloLens glasses and Unity game engine. For that purpose, we created an application that enables the interaction between user and six artefacts from Roman era, found in four different sites through the Balkan. In this article the advantages and disadvantages of such a presentation are presented and the possibility of its use for cultural heritage preservation.

Keywords

Cultural heritage, mixed reality, virtual reality, augmented reality, Microsoft HoloLens

1 Uvod

Kulturna dediščina je podedovana zapuščina, ohranjena v sedanjosti. V osnovi jo delimo na materialno in nematerialno. Z ohranjanjem kulturne dediščine se ukvarja UNESCO, agencija znotraj organizacije Združenih narodov, ki je na seznam kulturne dediščine uvrstila 1121 območij [1]. Od tega jih je kar 53 ogroženih zaradi naravnih katastrof, vremenskih sprememb, vojn in človeške malomarnosti. Zaradi razvoja in priljubljenosti sodobnih tehnologij, so njihovo vrednost prepoznali tudi pri ohranjanju kulturne dediščine. Več milijonov turistov, ki vsakoletno obiskujejo znamenitosti po svetu, bi lahko namesto dolgih potovanj in visokih stroškov iz udobja domačega fotelja doživeli cenejšo, ampak še vedno zadovoljivo izkušnjo, hkrati pa s tem tudi te znamenitosti razbremenili in tako pripomogli pri ohranjanju kulturne dediščine tudi za prihodnje generacije.

Problem izginjanja kulturne dediščine in priložnost njenega ohranjanja za prihodnje generacije s pomočjo sodobnih tehnologij smo želeli preveriti tudi v praksi. Izdelali smo aplikacijo za prikazovanje artefaktov iz časov Rimljanov, ki deluje na »pametnih« očalih za prikazovanje mešane resničnosti, Microsoft HoloLens in uporabniku omogoča osnovno interakcijo z artefakti ter njihov nemoten ogled iz vseh smeri.

Članek sestavlja pet razdelkov. V drugem razdelku predstavimo obstoječe aplikacije obogatene in mešane resničnosti, ki se uporabljajo pri ohranjanju kulturne dediščine. Naslednji razdelek, to je tretji, nas seznani z načrtovanjem in izdelavo aplikacije. V četrtem razdelku predstavimo delovanje aplikacije. Zadnji, peti razdelek je namenjen predstavitvi doseženih rezultatov.

2 Pregled področja

Čeprav sodobne tehnologije v javnosti pogosto označujejo kot grožnjo, ki lahko privede v odvisnost, socialno izolacijo in zmanjša ustvarjalnost [2], so ravno te tehnološke inovacije pogosto ključne pri ohranjanju kulturne dediščine za prihodne generacije. V zadnjih letih med najbolj priljubljene tehnologije za ohranjanje kulturne dediščine štejemo navidezno, obogateno in mešano resničnost.

S stališča računalništva štejemo pod pojem navidezna resničnost področje, katerega cilj je ustvariti virtualni svet, ki omogoča interakcijo z uporabnikom, medtem ko uporablja posebne naprave za simulacijo okolja, ki skrbijo za čim bolj realno izkušnjo. Obogatena resničnost se za razliko od navidezne osredotoča na dopolnitev resničnega sveta s pomočjo dodajanja plasti virtualnih objektov ali dodatnih informacij v resnično okolje.

Priljubljenost obogatene resničnosti z leti eksponentno narašča, vendar so v raziskavi ohranjanja kulturne dediščine na evropskem območju [3] ugotovili, da trenutno obstaja zelo malo aplikacij, ki jih večinoma razvijajo muzeji oziroma ustanove za ohranjanje kulturne dediščine [4]. Aplikacije se večinoma aktivirajo na podlagi sprožilca (npr. simbol, označba, predmet, lokacija naprave), v manjši meri pa tudi na podlagi pogleda.

Mobilna aplikacija England Originals [1] ter funkcija Pocket Gallery [2] znotraj aplikacije Google Arts&Culture [3] delujeta na podoben način – ob zaznavi ravne površine prikažeta tridimenzionalni model v resničnem okolju. Manipulacija z modelom je mogoča s premikanjem telefona ter upravljanjem preko zaslona. Tudi mobilna aplikacija Civilisations AR [4] deluje na enak način, vendar v tem primeru virtualni model našega planeta lebdi v zraku, na njem pa so označena najdišča artefaktov, ki si jih lahko s klikom na zaslon mobilne naprave tudi bolj natančno ogledamo ter nad njimi izvajamo osnovne geometrijske transformacije.

Aplikacije za ohranjanje kulturne dediščine s pomočjo obogatene resničnosti zelo uspešno vključujejo v turistično ponudbo tudi v Sloveniji, kjer lahko izpostavimo tri aplikacije. Travel AR Slovenia [5] sproži obogatitev okolice na prenosni napravi ob zaznavi markerja obogatene resničnosti in nam ob avdio vodenju omogoča ogled tridimenzionalnih rekonstrukcij kulturne dediščine Kočevja in Črnomlja. Na podoben način deluje tudi aplikacija AR Kranj [6], s pomočjo katere lahko spoznamo mesto Kranj in njegovo zgodovino. Kulturno dediščino lahko s pomočjo obogatene resničnosti spoznamo tudi v nekdanjem samostanu Žička Kartuzija, ki je danes v ruševinah, kjer s pomočjo pametnih očal in avdio vodnikov skozi tridimenzionalne modele vidimo, kako je samostan izgledal v vsej svoji veličini.

Kontinuum virtualnosti, katerega avtorja sta Paul Milgram in Fumio Kishino se osredotoča na definicijo mešane resničnosti. Po njuni definiciji mešano resničnost sestavljata obogatena resničnost, kjer virtualni elementi obogatijo resnični svet, ter obogatena virtualnost, kjer elementi resničnega sveta obogatijo virtualni svet. Tako lahko mešano resničnost predstavimo tudi kot nadmnožico obogatene in navidezne resničnosti.

Aplikacije mešane resničnosti se za ohranjanje kulturne dediščine v času pisanja tega članka uporabljajo zgolj v manjšem obsegu [5]. Možna razloga za to sta lahko visoka cena in relativno nova tehnologija.

HoloTour [7] je produkt podjetja Microsoft, ki nam omogoča ogled zgodovine Rima in skrivnosti Machu Picchua na napravi za prikazovanje mešane resničnosti, Microsoft HoloLens. Z aplikacijo upravljamo s pomočjo gest in glasovnih ukazov. Mešano resničnost vse pogosteje vključujejo tudi v muzeje, kjer lahko izpostavimo aplikaciji HoloMuse [8] in Holomuseum [9] (obe sta namenjeni za uporabo na napravi Microsoft HoloLens), ki uporabniku omogočata spoznavanje arheoloških zbirk artefaktov, s katerimi lahko poljubno manipuliramo, česar v pravem muzeju ne moremo doseči.

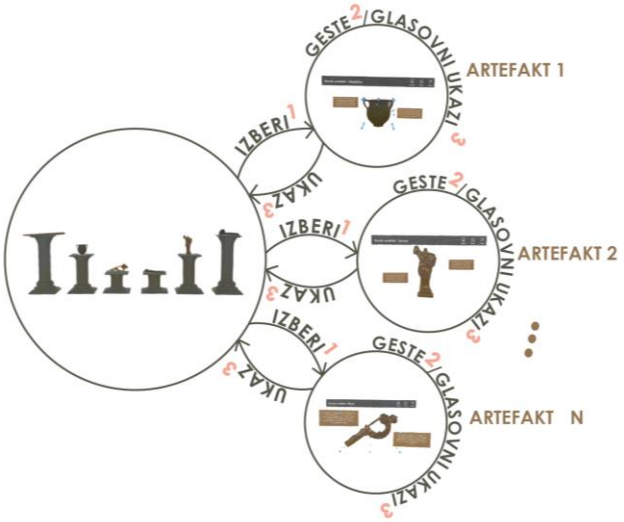
3 Načrtovanje in izdelava aplikacije

Pri snovanju aplikacije, nastajala je v času diplomskega dela, smo se zgledovali po obstoječih aplikacijah za prikazovanje mešane resničnosti. Pri tem pa smo se srečali s težavo, da gre pri očalih HoloLens za dokaj novo napravo, ki je cenovno težje dostopna. Ob tem je potrebno poudariti, da se knjižnice z novimi funkcionalnostmi še vedno razvijajo, zato je ustrezne literature dokaj malo, oziroma v njej avtorji opisujejo starejše verzije knjižnic, ki se več ne uporabljajo. Zato smo si zamislili enostavno, ampak še vedno vabljivo aplikacijo, skozi katero so predstavljene različne možnosti uporabe mešane resničnosti s pomočjo očal HoloLens.

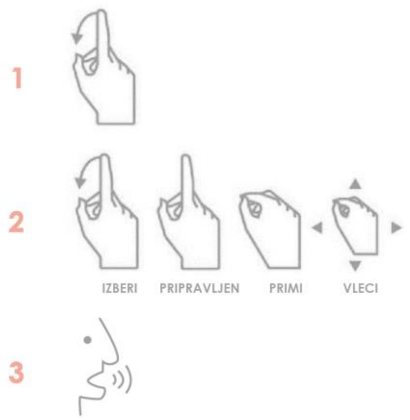
Predstavljena aplikacija spada pod obogateno resničnost, ki pa bi lahko postala aplikacija mešane resničnosti v primeru, da bi se pojavila potreba po tem, tako da bi uporabljenim objektom dodali zavedanje okolice. Aplikacija namreč deluje na napravi za

[1] http://www.heritagecities.com/stories/explore [09. 09. 2020].
[2] https://artsandculture.google.com/story/5QWhvYU1kBJfgw [09. 09. 2020].
[3] https://about.artsandculture.google.com/ [09. 09. 2020].
[4] https://www.bbc.com/news/technology-42966371 [09. 09. 2020].
[5] http://www.travel-ar.si/sl/ [09. 09. 2020].
[6] https://www.visitkranj.com/sl/obogatena-resnicnost-v-kranju [09. 09. 2020].
[7] https://docs.microsoft.com/en-us/windows/mixed-reality/case-study-capturing-and-creating-content-for-holotour [09. 09. 2020].
[8] https://www.researchgate.net/publication/315472858_HoloMuse_Enhancing_Engagement_with_Archaeological_Artifacts_through_Gesture-Based_Interaction_with_Holograms [09. 09. 2020].
[9] https://www.researchgate.net/publication/326713622_HOLOMUSEUM_A_HOLOLENS_APPLICATION_FOR_CREATING_EXTENSIBLE_AND_CUSTOMIZABLE_HOLOGRAPHIC_EXHIBITIONS [09. 09. 2020].

prikazovanje mešane resničnosti, Microsoft HoloLens, ki to v osnovi omogoča.



Slika 1: Shematski prikaz delovanja aplikacije



Slika 2: Osnovna načina interakcije z HoloLens, ki smo jih uporabili v aplikaciji (geste in glasovno upravljanje)

Aplikacijo smo razdelili na različne scene, kot prikazuje slika 1, med katerimi se lahko premikamo z uporabo osnovnih načinov interakcije s HoloLens, kot prikazuje slika 2.

3.1 Izdelava aplikacije

Prvi korak pri izdelavi aplikacije za razstavo muzejskih eksponatov s pomočjo mešane resničnosti, je priprava objektov. Osredje mesto na razstavi so zasedali eksponati s štirih nekdanjih rimskih naselbin: Viminacium in Municipium v Srbiji, Aquae v Bosni in Hercegovini ter Dyrrachium v Albaniji, ki smo jih dobili v elektronski obliki, vendar jih je bilo potrebno pred uporabo v naši aplikaciji obdelati. V ta namen smo uporabili animacijski paket Blender, v katerem smo zmanjšali število točk in mnogokotnikov, kot prikazuje tabela 1. Število točk je bilo potrebno zmanjšati zaradi procesorske moči in velikosti pomnilnika na napravi HoloLens.

Tabela 1: Število mnogokotnikov pred in po decimaciji

Objekt	Število mnogokotnikov pred decimacijo	Število mnogokotnikov po decimaciji
Artefakt 1	3.268.685	490.301
Artefakt 2	448.881	44.877
Artefakt 3	422.978	42.282
Artefakt 4	124.764	43.666
Artefakt 5	146.228	29.244
Artefakt 6	97.762	9.437

Za večjo atraktivnost aplikacije, smo dodatno zmodelirali še steber v slogu rimske arhitekture, na katere smo v predstavitvi nato postavili artefakte. Vse objekte smo nato izvozili v grafični pogon Unity.

V igralnem pogonu Unity smo ustvarili nov projekt ter namestili in konfigurirali orodje, ki omogoča razvoj aplikacij mešane in obogatene resničnosti, imenovano »Mixed Reality Toolkit«, oziroma krajše MRTK. Aplikacijo smo razdelili na posamezne dele imenovane scene, ki smo jih uporabili kot samostojne enote. V vsako izmed scen smo dodali izbrane objekte ter jim dodali interaktivne komponente, ki so nam omogočile interakcijo z artefakti. Interakcija s HoloLens je možna na tri osnovne načine: s pogledom, kretnjo ali glasovnim upravljanjem. Vse tri načine smo implementirali v našo aplikacijo.

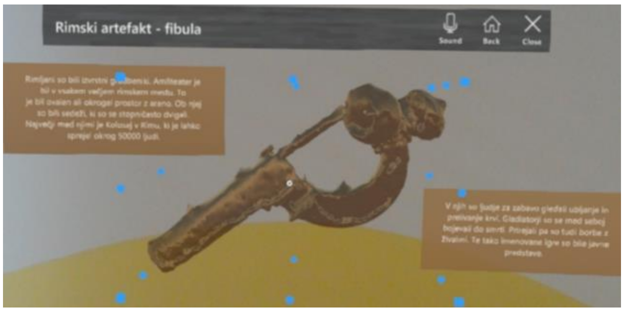
4 Predstavitev rezultatov

Aplikacijo za ohranjanje kulturne dediščine s pomočjo navidezne in obogatene resničnosti smo razvili skozi različne stopnje, ki skupaj sestavljajo celoto; delujočo aplikacijo na napravi za prikaz mešane resničnosti Microsoft HoloLens, skozi katero lahko spoznamo artefakte iz časov Rimljanov.

Ob njenem zagonu se nam prikaže glavna scena, ki prikazuje šest stebrov, kot prikazuje slika 3. Na vsakega izmed stebrov je postavljen po en artefakt iz rimske dobe. Za izboljšano uporabniško izkušnjo izleta v preteklost, se v ozadju predvaja glasba s časa Rimskega imperija, kot si jo danes predstavljajo muzikologi.

Slika 3: Glavna scena s šestimi artefakti, ki so postavljeni na različno visoke stebre

Artefaktov pa si ni možno ogledovati zgolj od daleč, ampak si jih lahko ogledamo tudi natančneje. Za dostop do scene posameznega artefakta usmerimo pogled v artefakt ter z gesto, ki ponazarja klik, preidemo na novo sceno, kot prikazuje slika 4.



Slika 4: Prikaz posamezne scene enega izmed artefaktov

Na tej sceni lahko nad artefaktom izvajamo osnovne geometrijske transformacije (spreminjanje velikosti, rotiranje in premikanje) s pomočjo prijemanja ročajev ob straneh artefakta. Na takšen način lahko artefakt pogledamo iz vseh smeri, česar v realnem muzeju ne moremo doseči. Ob straneh artefakta se prikažeta dve ploščici, na katerih sta zapisani zanimivosti o življenju Rimljanov. Po želji jih lahko z usmeritvijo pogleda in gesto, ki ponazarja klik, zapremo in vso pozornost usmerimo v artefakt.

Na sceni, ki prikazuje artefakte posamezno, se bo nad njimi pojavila nadzorna plošča, ki omogoča prekinitev predvajanja glasbe, vrnitev v glavno sceno; po želji pa lahko kontrolno ploščo tudi zapremo. Vse naštete ukaze izvedemo z usmeritvijo pogleda v izbrano akcijo (gumb) in gesto, ki ponazarja klik. Ob usmeritvi pogleda v željeni gumb se nam pod njim izpiše ključna beseda, s katero lahko izvedemo ukaz.
Sceno, ki prikazuje posamezen artefakt, lahko upravljamo tudi s štirimi glasovnimi ukazi. Prvi ukaz s ključno besedo »Menu« se uporablja v primeru, da smo pred tem nadzorno ploščo zaprli. Nadzorna plošča se bo ob zaznavi ukaza ponovno prikazala. Drugi ukaz s ključno besedo »Sound« se uporablja v primeru, ko smo pred tem predvajanje glasbe ustavili. Ob zaznavi ukaza se bo glasba predvajala naprej. Tretji ukaz s ključno besedo »Close« se uporablja kot nadomestilo klika na gumb Close in nam ob zaznavi ukaza zapre nadzorno ploščo. Zadnji, četrti ukaz, s ključno besedo »Back«, se uporablja kot nadomestek klika na gumb Back in nam ob zaznavi ukaza ponovno prikaže glavno sceno.

Aplikacijo smo želeli tudi testirati na testnih uporabnikih in pridobiti njihov odziv, vendar to zaradi epidemiološke situacije glede COVID-19 ni bilo mogoče. Pričakujemo, da bi bili rezultati podobni rezultatom, predstavljenim v članku [6], saj avtorji tam med drugim opisujejo uporabniške izkušnje s podobno aplikacijo kot je naša.

5 Zaključek

V tem projektu smo izpostavili problem izginjanja kulturne dediščine, ki je posledica različnih dejavnikov (naravnih in človeških) ter želeli preveriti možnosti ohranjanja naše preteklosti s pomočjo sodobne tehnologije, mešane resničnosti. Razvili smo aplikacijo, ki nam omogoča ogled artefaktov iz časov Rimljanov s pomočjo naprave za prikazovanje mešane resničnosti Microsoft HoloLens.

Čeprav smo uspešno razvili aplikacijo, kot smo si jo zamislili, smo mnenja, da imata tako aplikacija kot tehnologija mešane ter obogatene resničnosti veliko možnost nadgradnje v prihodnosti. Največjo omejitev pri razvoju trenutno predstavlja naprava HoloLens in njene tehnične zmogljivosti, kot je npr. slabša ločljivost in majhno vidno polje na vizirju. Druga generacija Microsoftove naprave bi naj izboljšala vse slabosti naprave prejšnje generacije in omogočila bolj naravno interakcijo s hologrami.

Če smo v preteklosti razmišljali, kako prilagoditi objekte, da bodo lahko sprejeli več obiskovalcev, bomo v prihodnosti morali več pozornosti nameniti uporabi sodobnih tehnologij na različnih področjih, tudi ohranjanju kulturne dediščine. Obogatena in mešana resničnost sta vsekakor tehnologiji, ki ju lahko uporabimo na kateremkoli področju. Potrebujemo zgolj tridimenzionalne modele in zgodbo, ki bo pritegnila uporabnike, zgodbo, ki je sestavni del naše preteklosti, preteklosti, ki jo želimo ohraniti za prihodnje generacije.

Zahvala

Avtorji izjavljamo, da je raziskavo finančno podprla Javna agencija za raziskovalno dejavnost Republike Slovenija v okviru projekta BI-BA/19-20-003.

6 Viri in literatura

[1] Unesco. World Heritage List. Dostopno na: https://whc.unesco.org/en/list/ [26.2.2020].

[2] Al-Zoubi, S., Younes, M. A. B. The Impact of Technologies on Society: A Review. *International Organization of Scientific Research Journal of Humanities and Social Science*, 20, (2015), 2(5), str. 82-86.

[3] Luna, U., Rivero, P., Vicent, N. Augmented Reality in Heritage Apps : Current Trends in Europe. *Applied Sciences*, 9, (2019), 13, 2756.

[4] Hammady, R., Ma, M., Strathearn, C. User experience design for mixed reality: a case study of HoloLens in museum. *International Journal of Technology Marketing*, 13, (2019), 3/4, str. 354-375.

[5] Kassahun-Bekele M., Pierdicca, R., Frontoni, E., Malinverni, E. S., Gain, J. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *Association for Computing Machinery Journal on Computing and Cultural Heritage*, 11, (2018), 2, str. 1-36.

[6] Kolmanič, S., Marksel, M., Mongus, D., Žalik, B. Tehnologije navidezne in obogatene resničnosti, kot orodje za predstavitev novih idej in produktov na sejmih : primer Mahepa. *Uporabna informatika*. [Tiskana izd.]. 2020, letn. 28, št. 2, str. 85-93. ISSN 1318-1882.

Predmetnik: oprijemljiv uporabniški vmesnik za informiranje turistov

Gregor Sotlar
89172027@student.upr.si
Univerza na Primorskem, UP FAMNIT
Koper, Slovenija

Peter Roglej
peter.rogelj@upr.si
Univerza na Primorskem, UP FAMNIT
Koper, Slovenija

Klen Čopič Pucihar
klen.copic@famnit.upr.si
Univerza na Primorskem, UP FAMNIT
Koper, Slovenija
Fakulteta za Informacijske Študije
Novo mesto, Slovenija

Matjaž Kljun
matjaz.kljun@famnit.upr.si
Univerza na Primorskem, UP FAMNIT
Koper, Slovenija
Fakulteta za Informacijske Študije
Novo mesto, Slovenija

Slika 1: Vmesnik Predmetnika.

POVZETEK

Namen dela je zasnovati in raziskati možne vloge oprijemljivega uporabniškega vmesnika za informiranje turistov, ki bi dopolnjeval obstoječe oblike informiranja v turistično informacijskih centrih, nadomestil njihove pomanjkljivosti, a hkrati uporabil njihove prednosti. Vmesnik smo zasnovali na podlagi predhodnih raziskav in lastnih izkušenj ter ga poimenovali Predmetnik. Uporabniški vmesnik vsebuje posamezne enote - predmete, ki predstavljajo določeno turistično ponudbo. Ob dvigu predmeta uporabnik sproži prikaz vsebine na zaslonu. Opravljena uporabniška študija je pokazala, da je lahko Predmetnik prva točka informiranja v turistično informacijskih centrih, njegova prednost pa je v tem, da na enostaven in preprost način v kratkem času podaja informacije o doživetju posamezne turistične ponudbe.

KEYWORDS

oprijemljivi uporabniški vmesnik, turizem, informiranje turistov, turistični informacijski center, TIC

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia

© 2020 Copyright held by the owner/author(s).

1 UVOD

Informiranje turistov v turistično informacijskih centrih (TIC) je omejeno na nekaj medijev ali virov, kot so turistični informator, tiskovine (letaki, prospekti, zemljevidi, brošure in zgibanke), zasloni, ki predvajajo videoposnetke, in računalniki (npr. zasloni na dotik). Težave tiskanih medijev so enosmerna komunikacija, težko posodabljanje, različna grafična razporeditev, poleg tega težko omogočajo nadaljnje iskanje informacij (npr. preko URL naslovov, ki jih je potrebno prepisati v brskalnik). Tiskovine imajo tudi prednosti, saj ne potrebujejo dodatne energije in so zelo prenosljive.

Težave zaslonov na dotik so zahtevano znanje za uporabo, čas, ki ga porabimo pri iskanju informacij, potrebno napajanje z energijo in nezmožnost odnesti informacije s seboj. Podobno kot tiskani mediji tudi neinteraktivni zasloni (angl. *public displays*) omogočajo povečini enosmerno komunikacijo s predvajanjem videoposnetkov turistične ponudbe. Ti so zelo dobrodošli za prvi vtis o ponudbi, vendar teh informacij obiskovalec ne more upravljati in lahko le čaka, da se vse informacije predvajajo.

Za raziskavo informiranja o turistični ponudbi smo zasnovali in izdelali oprijemljiv uporabniški vmesnik, imenovan Predmetnik. Sistem odpravlja težavno interakcijo z zasloni na dotik in nezmožnost interakcije z javnimi zasloni. Vmesnik je sestavljen iz predmetov, s katerimi lahko uporabnik rokuje in preko tega upravlja večpredstavnostne vsebine o določenem turističnem cilju, aktivnosti ali ponudbi na povezanem zaslonu. Za predmete našega vmesnika lahko izberemo lokalne izdelke, pridelke, spominke in različne predmete, ki so povezani z aktivnostmi. Predmeti, s katerimi uporabnik rokuje, preko zaslona predstavljajo določeno zgodbo, ki obogati turistično izkušnjo. Predmeti s tem postanejo "vstopna točka" in preko podanih zgodb spodbudijo željo po iskanju nadaljnjih informacij, ki so na voljo v TIC-u, kot na primer, kako priti do želene posamezne ponudbe ali cilja, zgodovino, ipd. Vse dodatne informacije so torej dosegljive preko tiskovin, spleta in turističnih informatirjev, ki so že na voljo v TIC-u.

2 PREGLED PODROČJA

Skupnost je že pred tremi desetletji izpostavila, da računalniki preprečujejo stik z okoljem [26], kar je spodbudilo ideje za uporavljanje digitalnih vsebin s pomočjo fizičnih predmetov [7]. Nekateri raziskovalci so šli še dlje in predstavili vizijo uporabe fizičnega sveta kot vmesnika za povezovanje objektov in površin z digitalnimi vsebinami [12]. Na osnovi teh del so oprijemljivi uporabniški vmesniki postali nova oblika interakcije [18], ki se uporablja na vse več področijih in za raznovrstne naloge [21], kot so: (i) shranjevanje, pridobivanje in rokovanje s podatki [1, 5, 19, 22], (ii) vizualizacija informacij preko oprijemljivih uporabniških vmesnikov [10, 23, 24], (iii) modeliranje in simulacije [2–4, 8, 11], (iv) upravljanje sistemov, kontrola in konfiguracija [3, 5, 14, 15, 22] in (v) izobraževanje, zabava in programski sistemi [3, 9, 13, 16, 17].

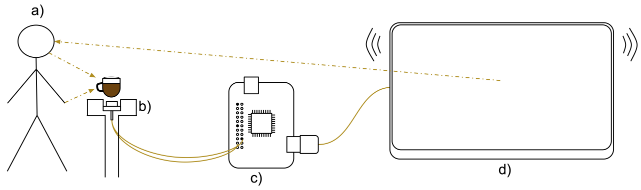
Tudi na področju turizma so že bile narejene raziskave, ki so za informiranje turistov uporabile koncepte oprijemljivih uporabniških vmesnikov. Sistem *Mementos* [6] je uporabnike preko žetonov (spominkov), ki so predtsavljali turistične znamenitosti ali infrastrukturo (restavracije, javni prevoz, ipd.), na po mestu postavljenih kiosk računalnikih, ki so žetone prepoznali, vodil do želenih ciljev. *Tangible user Interface within Projector-based Mixed Reality* je s sledenjem figure, ki jo je uporabnik premikal po zemljevidu, na zaslonu prikazoval pripadajoči 3D prizor [27]. Pri drugi izvedbi je fizično figuro zamenjal uporabnik sam, ki je s premikanjem po na tleh projicirani maketi parka, upravljal prikaz lokacije. Oprijemljivi vmesniki so pogosto dostopni v muzejih. Na primer v [25] je oprijemljiv vmesnik za pridobitev informacij o določenem geološkem vzorcu kar vzorec sam, ki ob rokovanju poda zvočne in vizualne informacije o njem na projekciji. Vzorec je hkrati tudi bogat vir informacij o barvi, teži, trdoti in teksturi.

Dosedanje raziskave so se osredotočale na prikaz poti do želene turistične znamenitosti, prikaz pokrajine glede na položaj in rokovanje s predmeti za prikaz podrobnejših (video) vsebin. Nobena nam znana raziskava se ni ukvarjala s podajanjem predstave o turističnih doživetjih preko različnih predmetov kot so spominki, lokalni pridelki ali izdelki in predmeti, ki predstavljajo turistične aktivnosti (pohodi, kolesarjenje, ipd.) Osnovna ideja je tako omogočiti upravljanje in izbiro aktivnosti ali ciljev potovanja iz predhodnega nabora fizičnih predmetov, ki predstavljajo asociacijo na določeno turistično izkušnjo, s tem pa omogočiti pridobitev možnega doživetja turistične ponudbe preko večpredstavnostnih vsebin, ki so drugače dosegljive le na spletu in javnih zaslonih. Naše raziskovalno vprašanje se tako glasi: Ali je fizični vmesnik Predmetnik primeren kot vstopna točka za informiranje o turističnih doživetjih oziroma o ponudbi v turistično informacijskih centrih?

3 OPIS SISTEMA

Predmetnik je sestavljen iz treh delov (Slika 2): uporabniškega vmesnika, mikroračunalnika (Raspberry Pi 3B) kot računske enote in naprave za predvajanje zvočnih in video vsebin (zaslon ali projektor). Komunikacija poteka v smeri od oprijemljivega vmesnika preko senzorjev do mikroračunalnika, ki prejme informacijo o dvignjenem predmetu, predvaja temu primerno vsebino in preko svetlobnih signalov podaja informacije o aktivnih in neaktivnih predmetih vmesnika (če so vsi predmeti odloženi, pri vseh gori lučka; če je pa posamezen predmet dvignjen, gori lučka le pri tem). Polica predmetnika ima za posamezen predmet izrezan relief v obliki predmeta, kar omogoča (poleg svetlobnega signala) lažje odlaganje. Če noben predmet ni aktiven (se z njim ne rokuje), se na zaslonu predvaja kratek video posnetek, ki prikazuje rokovanje s Predmetnikom in vabi uporabnika, da dvigne enega od njih. Podrobnejši opis uporabljene strojne in programke opreme je na voljo v [20].

Sistem zbira naslednje podatke: število rokovanj z določenim predmetom, čas rokovanja s posameznim predmetom in čas predvajanja posameznega posnetka s čimer beležimo,



Slika 2: Prikaz sheme sistema: a) uporabnik, b) uporabniški vmesnik s predmeti, c) računska enota in d) zaslon, ki predvaja večpredstavnostne vsebine.

koliko je bila posamezna ponudba zanimiva in koliko je bil določen predmet vmesnika informativen ali zanimiv za interakcijo.

4 RAZISKAVA

V raziskavi smo izvedli nadzorovano uporabniško študijo (angl. *controlled user study*). Opazovalne študije (angl. *observational study*) nismo mogli izvesti zaradi epidemije. Na TIC-u v Izoli so vse predmete odmaknili, saj jih turisti zaradi možnosti okužbe ne smejo prijemati; lahko jih dobijo le, če vprašajo zaposlenega informatorja. Zaradi tega tudi niso dovolili postaviti Predmetnik v njihove prostore.

Slika 3: Simulacija TIC-a: zaslon na dotik, Predmetnik in tiskovine.

Študijo smo izvedli v simuliranem TIC-u (Slika 3). Na voljo so bile tri oblike informacij, ki so predstavljale turistično ponudbo slovenske Istre: tiskovine, zaslon na dotik (tablica s spletno stranjo *I Feel Slovenia: Mediterranean & Karst Slovenia*[1]) in Predmetnik (s štirimi predmeti – pedal, vponka, sol in kamen s pohodniško markacijo), ki je prikazoval video vsebine s spletnih strani turističnih zavodov Kopra, Izole in Pirana. Video vsebine so bile dolge med 12 in 19 sekund.

Udeležence smo pridobili s priročnim vzorčenjem (angl. *convenience sampling*). V danem trenutku in položaju je bila to edina možnost pridobitve uporabnikov. Pri testiranju je sodelovalo devet udeležencev v razponu od devet do 66 let, od tega je bilo sedem žensk in dva moška.

Po privolitvi smo jim predstavili potek raziskave: (i) izpolnjevanje pred-vprašalnika, (ii) opravljanje dveh nalog (iskanje informacij o kolesarjenju in pregled ostalih možnih doživetij) brez časovnih omejitev, kjer so imeli na voljo tiskovine, zaslon na dotik in Predmetnik ter (iii) izpolnjevanje drugega dela vprašalnika. Vrstni red opravljanja nalog ni bil naključen, saj smo najprej želeli videti rokovanje za točno določen namen (najti informacije o kolesarjenju) in nato opazovati splošno rokovanje (pregled možnih doživetij).

Med raziskovanjem smo opazovali interakcijo z vsako od oblik informiranja (obračanje, tipanje ...) Opazovali smo tudi vrstni red interakcije med raznimi oblikami oziroma predmeti informiranja, porabljen čas na posameznem viru interakcije, čas gledanja pri video predstavitvah in skupen čas iskanja informacij.

5 REZULTATI IN RAZPRAVA

V povprečju so udeleženci prvo nalogo reševali osem minut in 16 sekund, drugo pa pet minut in 31 sekund, čeprav je slednja od njih zahtevala pregled več informacij. Tiskovinam so pri drugi nalogi v povprečju namenili dobro minuto in pol manj časa v primerjavi s prvo nalogo, posamezna tiskovina pa se je v povprečju gledala enako dolgo, čeprav niso vsi iz prve naloge po tiskovinah posegli tudi v drugi. Pri tablici se je število uporabnikov pri drugi nalogi v primerjavi s prvo zmanjšalo, predvsem zaradi začetne slabe izkušnje z njenim rokovanjem. Ravno tako se je zmanjšal čas povprečne porabe za skoraj dve minuti. Le pri Predmetniku se je čas uporabe podaljšal iz 32 sekund na 53 sekund. To je bilo za pričakovati, saj so imeli možnost pogledati še tri preostale posnetke. Čas se je podaljašal tudi za ogled posameznega videoposnetka v povprečju za tri sekunde, vendar k temu prispeva tudi dejstvo, da so drugi trije videi za nekaj sekund daljši od videa, predvidenega za prvo nalogo. Povečalo se je zanimanje za Predmetnik, saj ga je po tem, ko so ga v prvi nalogi šele spoznali, v drugi nalogi kot prvo obliko informiranja izbralo več uporabnikov. Vrstni red nalog je deloma tudi vplival na čas opravljanja: slabe izkušnje iz prve naloge so vplivale na neuporabo tablice, domačnost s Predmetnikom pa na opustitev uvodnega videa.

Poleg tega se je v drugi nalogi v primerjavi s prvo povečalo število uporabnikov, ki so si začeli ogledovati predmete ali se z njimi igrati, vendar je bilo število manjše od pričakovanega. Previdevamo, da imajo tiskovine in tablica prednost, ker so jih udeleženci že poznali ali vsaj vedeli, kakšna je njihova funkcionalnost. Prišlo je tudi do določenih sprememb pri izvajanju obeh nalog, kot je povečanje prehajanja med oblikami informiranja oziroma vračanja k že obiskani obliki znotraj iste naloge. Še posebno pri prvi nalogi smo opazili, da so udeleženci najprej uporabili tisto obliko informiranja, ki

[1]https://www.slovenia.info/en/places-to-go/regions/mediterranean-karst-slovenia

jim je bila fizično najbližje. Tako so pri prvi nalogi večinoma začeli z uporabo tiskovin, ki so bile prva oblika informiranja glede na smer prihoda v prostor. Pri drugi nalogi pa je to že bil Predmetnik.

Predmetnik, ne nadomešča drugih virov saj sta namen in učinek različnih načinov pridobivanja informacij različen. (i) Predmetnik je privlačen in nudi maloštevilno izbiro informacij brez podrobnejšega usmerjanja znotraj področja zanimanja, (ii) tablica (zasloni na dotik) je manj privlačna a omogoča bolj podrobno usmerjanje k željenim informacijam z več nivojsko izbiro, (iii) tiskovine pa so edine izmed treh, ki jih obiskovalci lahko vzamejo s seboj in informacije iz njih prebirajo tudi kasneje, a ne nudijo večpredstavnostnih vsebin. Rezultate bi težko primerjali z drugimi raziskavami saj niso dovolj sorodne. Rezultati vprašalnikov so predstavljeni v [20].

6 ZAKLJUČEK

V članku je predstavljen oprijemljivi uporabniški vmesnik Predmetnik, ki bi turistom v TIC na hiter in enostaven način prikazal doživljajske informacije kot osnovo za raziskovanje turistične ponudbe. Vmesnik smo izdelali kot nadgradnjo obstoječih zaslonov na dotik in javnih zaslonov (angl. *public displays*), ki le prikazujejo videoposnetke. Namen vmesnika je tako postati vstopna točka informiranja za turiste, ki bi lahko preko tiskovin, turističnega informatorja in spleta nato naprej raziskovali turistično ponudbo, ki bi jih na Predmetniku pritegnila.

S Predmetnikom smo izvedli nadzorovano uporabniško študijo. Povečan čas uporabe in večkratnega rokovanja s predmeti pri drugi nalogi deloma odgovori na zastavljeno vprašanje o vlogi Predmetnika, ki lahko predstavlja vstopno točko pri informiranju, vendar bi bilo v prihodnosti treba izvesti obširnejšo študijo v realnem okolju TIC-a s turisti, kar v danem trenutku zaradi pandemije ni bilo mogoče. Poleg tega bi bilo potrebno razširiti ponudbo Predmetnika ter dodati druge funcionalnosti (ponujanje nadaljnjega raziskovanja (QR kode, povezovanje z drugimi viri v enotno izkušnjo), sledenje pogledu, priporočilni sistem, sledenje predmetom ...). Bolj podrobno je vse opisano v [20].

LITERATURA

[1] Rachel Abrams. 1999. Adventures in tangible computing: The work of interaction designer 'Durrell Bishop'in context. *Master's thesis, Royal College of Art, London* (1999).

[2] Robert Aish. 1979. 3D input for CAAD systems. *Computer-Aided Design* 11, 2 (1979), 66–70.

[3] George Anagnostou, Daniel Dewey, and Anthony T. Patera. 1989. Geometry-defining processors for engineering design and analysis. *The Visual Computer* 5, 5 (1989), 304–315.

[4] David Anderson, James L Frankel, Joe Marks, Aseem Agarwala, Paul Beardsley, Jessica Hodgins, Darren Leigh, Kathy Ryall, Eddie Sullivan, and Jonathan S Yedidia. 2000. Tangible interaction+ graphical interpretation: a new approach to 3D modeling. , 393–402 pages.

[5] Jonathan Cohen, Meg Withgott, and Philippe Piernot. 1999. Logjam: a tangible multi-person interface for video logging. , 128–135 pages.

[6] Augusto Esteves and Ian Oakley. 2010. Mementos: A Tangible Interface Supporting Travel. , 4 pages. https://doi.org/10.1145/1868914.1868994

[7] George W Fitzmaurice, Hiroshi Ishii, and William AS Buxton. 1995. Bricks: laying the foundations for graspable user interfaces. , 442–449 pages.

[8] John Frazer. 1995. An evolutionary architecture. (1995).

[9] Matthew G Gorbet, Maggie Orth, and Hiroshi Ishii. 1998. Triangles: tangible interface for manipulation and exploration of digital information topography. , 49–56 pages.

[10] Ken Hinckley, Randy Pausch, John C Goble, and Neal F Kassell. 1994. Passive real-world interface props for neurosurgical visualization. , 452–458 pages.

[11] Hiroshi Ishii, Carlo Ratti, Ben Piper, Yao Wang, Assaf Biderman, and Eran Ben-Joseph. 2004. Bringing clay and sand into digital design—continuous tangible user interfaces. *BT technology journal* 22, 4 (2004), 287–299.

[12] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. , 234–241 pages.

[13] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The ReacTable: Exploring the Synergy between Live Music Performance and Tabletop Tangible Interfaces. , 8 pages. https://doi.org/10.1145/1226969.1226998

[14] Jeevan James Kalanithi and V Michael Bove Jr. 2015. Tangible social network. US Patent 9,002,752.

[15] Martin Kaltenbrunner, Till Bovermann, Ross Bencina, Enrico Costanza, et al. 2005. TUIO: A protocol for table-top tangible user interfaces. , 5 pages.

[16] Radia Perlman. 1976. Using computer technology to provide a creative learning environment for preschool children. (1976).

[17] Hayes Solos Raffle, Amanda J Parkes, and Hiroshi Ishii. 2004. Topobo: a constructive assembly system with kinetic memory. , 647–654 pages.

[18] O. Shaer and E. Hornecker. 2010. *Tangible User Interfaces: Past, Present, and Future Directions.* NOW.

[19] Andrew Singer, Debby Hindus, Lisa Stifelman, and Sean White. 1999. Tangible progress: less is more in Somewire audio spaces. , 104–111 pages.

[20] Gregor Sotlar. 2020. *Oprijemljiv uporabniški vmesnik za informiranje turistov.* Master's thesis. Univerza na Primorskem, Titov trg 4, 6000 Koper.

[21] Brygg Ullmer and Hiroshi Ishii. 2000. Emerging frameworks for tangible user interfaces. *IBM systems journal* 39, 3.4 (2000), 915–931.

[22] Brygg Ullmer, Hiroshi Ishii, and Dylan Glas. 1998. mediaBlocks: physical containers, transports, and controls for online media. , 379–386 pages.

[23] Brygg Ullmer, Hiroshi Ishii, and Robert JK Jacob. 2003. Tangible query interfaces: Physically constrained tokens for manipulating database queries. , 279–286 pages.

[24] John Underkoffler and Hiroshi Ishii. 1999. Urp: a luminous-tangible workbench for urban planning and design. , 386–393 pages.

[25] Roberto Ivo Fernandes Vaz, Paula Odete Fernandes, and Ana Cecília Rocha Veiga. 2016. Proposal of a tangible user interface to enhance accessibility in geological exhibitions and the experience of museum visitors. *Procedia Computer Science* 100 (2016), 832–839.

[26] Pierre Wellner, Wendy Mackay, and Rich Gold. 1993. Back to the Real World. *Commun. ACM* 36, 7 (July 1993), 24–26. https://doi.org/10.1145/159544.159555

[27] Yuan Yuan, Xubo Yang, and Xiao Shuangjiu. 2007. A Framework for Tangible User Interfaces within Projector-based Mixed Reality. , 283-284 pages.

Razvoj in Ocenjevanje Prototipa Mobilne Aplikacije z Elementi Igrifikacije in Mešane Resničnosti

Development and Assessment of the Mobile Application Prototype with Elements of Gamification and Mixed Reality

Monika Zorko[†]
Fakulteta za elektrotehniko, računalništvo in informatiko
Univerza v Mariboru
Slovenija
monika.zorko1@student.um.si

Matjaž Debevc
Fakulteta za elektrotehniko, računalništvo in informatiko
Univerza v Mariboru
Slovenija
matjaz.debevc@um.si

Ines Kožuh
Fakulteta za elektrotehniko, računalništvo in informatiko
Univerza v Mariboru
Slovenija
ines.kozuh@um.si

ABSTRACT / POVZETEK

Učinkovito oglaševanje je eden od ključnih ciljev snovalcev oglasov in njihovih naročnikov. Prav zato si prizadevajo, da njihovi oglasi izstopajo od konkurence, pogosto pa je pri tem spregledan vidik uporabnika. V raziskovalni študiji smo tako izdelali prototip aplikacije, ki vključuje elemente igrifikacije in mešane resničnosti. Zaradi omejitev osebnih stikov v času pandemije COVID-19 smo izdelali video posnetke, ki so prikazovali uporabo prototipa. Nato smo ocenjevali uporabniško izkušnjo prototipa in uporabnost uporabniškega vmesnika. Uporabili smo SUS in UEQ metodo. S priložnostnim vzorčenjem smo v raziskavo vključili 80 oseb. Statistična analiza je razkrila tri ključne ugotovitve. Oglas, ki vsebuje mešano resničnost in igrifikacijo, nekoliko izstopa od ostalega načina oglaševanja. Prav tako lahko taka vrsta oglasa poveča stopnjo namena nakupa oglaševanega izdelka. Kot zadnje se je pokazalo, da ni povezave med starostjo uporabnika in razumevanjem aplikacije. Naši rezultati lahko služijo tako oglaševalcem, kot tudi raziskovalcem na področju uporabe sodobnih tehnologij in oglaševanja.

KEYWORDS / KLJUČNE BESEDE

uporabniška izkušnja, uporabnost, oglaševanje, igrifikacija, mešana resničnost

OPTIONAL: ABSTRACT

Effective advertising is one of the key goals of ad creators and their target groups. This is why they strive to make their ads stand out from the competition, while the user aspect is regularly overlooked. In the current study, we thus produced a prototype application that includes elements of gamification and mixed reality. Due to the limitations of personal contact during the COVID-19 pandemic, we produced videos showing the use of the prototype. We then evaluated the user experience of the prototype and the usability of the user interface. We used the SUS and UEQ method. 80 people were included in the survey by random sampling. Statistical analysis revealed three key findings. An ad that contains mixed reality and gamification stands out slightly from the rest of the advertising method. This type of ad can also increase the level of intent to purchase the advertised product. Lastly, the analysis revealed that there is no association between users' age and the understanding of the application. Our results can serve both advertisers and researchers in the use of modern technologies and advertising.

OPTIONAL: KEYWORDS

User experience, usability, advertising, gamification, mixed reality

1 UVOD

Vsakodnevno smo izpostavljeni številnim oglasom, kar vodi oglaševalce v vse večja vlaganja v zagotavljanje učinkovitosti oglaševanja in razlikovanja od konkurence. Sodobna tehnologija daje oglaševalcem številne možnosti za inovativne pristope v komuniciranju s ciljnimi javnostmi. Primera takih pristopov sta vpeljava igrifikacije in mešane resničnosti v oglaševanje. Oboje se je izkazalo kot pozitiven dejavnik v priklicu blagovne znamke s strani potrošnika [1].

Namen pričujoče študije je tako raziskati neizkoriščen potencial, ki ga prinaša oglaševanje s pomočjo kombinacije igrifikacije in mešane resničnosti. Natančneje, zanima nas zaznana stopnja vidljivosti oglasa, ki vpeljuje igrifikacijo in mešano resničnost v zgodbo komuniciranja s potrošnikom. Prav tako raziskujemo vplive na odločitve za nakup s tovrstnimi oglasi oglaševanih izdelkov. Ker se v procesu oblikovanja tovrstnih oglasov pojavljajo tudi izzivi v smislu zagotavljanja ustrezne uporabniške izkušnje in uporabnosti uporabniškega vmesnika, je predmet te študije raziskati tudi to.

2 IGRIFIKACIJA, RAZŠIRJENA RESNIČNOST IN OGLAŠEVANJE

Pri igrifikaciji gre za uporabo izkušnje zabave, ki »z notranjo motivacijo in sistemom nagrajevanja uporabnike privlači in jih vključi v različne aktivnosti« [2]. Tipični elementi igrifikacije so

*Article Title Footnote needs to be captured as Title Note
†Author Footnote to be captured as Author Note

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

točke, značke, lestvice, grafi uspešnosti, zgodbe s pomenom, avatarji in soigralci [4].

Mešana resničnost je del razširjene resničnosti, kamor jo uvrščamo skupaj z navidezno resničnostjo in obogateno resničnostjo. Mešano resničnost lahko opišemo kot tehnologijo med povsem resničnim okoljem in povsem navideznim okoljem [5]. Gre za okolje, kjer sta navidezni in resnični svet v enem samem zaslonu. Za njeno delovanje se uporablja dovolj zmogljiva tehnologija, kar zajema ustrezen senzor, procesor in zaslon.

3 PREGLED SORODNIH DEL

Obstoječe raziskave se intenzivno ukvarjajo z vprašanjem učinkovitosti vpeljave igrifikacije kot inovativnega orodja v oglaševanje [6]-[8]. In sicer, Nobre in Ferreira [6] v svoji študiji ugotavljata, da je s pomočjo igrifikacije mogoče na inovativen način soustvarjati blagovno znamko, vplivati na vpletenost uporabnika in občutek povezanosti z blagovno znamko. Teotónio in Reis [7] ugotavljata, da porabniki iščejo zabavo, nagrade, rivalstvo, socialno vključenost – vse, kar jim ponuja igrifikacija.

Prav tako se številni raziskovalci [9][10] ukvarjajo z elementi obogatene resničnosti v oglaševanju. Tako ugotavljajo, da se z uporabo tovrstne aplikacije poveča interakcija kupca in prodajalca, zviša ugled podjetja ter nenazadnje poviša tudi prodaja izdelkov [3],[10].

4 METODOLOGIJA

4.1 Raziskovalna vprašanja

Raziskovalna vprašanja smo oblikovali na osnovi pregleda obstoječe literature.

RV1: Kakšna je zaznana stopnja vidljivosti oglasa s hkratno vpeljavo igrifikacije in mešane resničnosti v mobilno aplikacijo za oglaševanje?

Prvo raziskovalno vprašanje smo zastavili, saj se je vpeljava interaktivne igre v oglaševanje izkazala kot učinkovita in za uporabnika zanimiva metoda oglaševanja [8].

RV2: Kakšna je zaznana stopnja odločitve za nakup izdelka ob hkratni vpeljavi igrifikacije in mešane resničnosti v mobilno aplikacijo za oglaševanje?

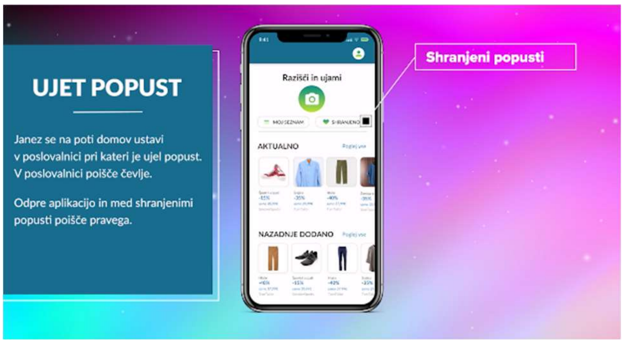
Drugo raziskovalno vprašanje smo zastavili, saj so pretekle raziskave pokazale, da se z uporabo mešane resničnosti lahko poveča stopnja prenosa informacij, sodelovanja ter pospeši odločanje [9]. Prav tako raziskave kažejo, da se z uporabo aplikacije, ki vsebuje obogateno resničnost, poveča interakcija kupca in prodajalca, zviša ugled podjetja in tudi poveča prodaja izdelkov [10].

RV3: Ali starost uporabnika vpliva na razumevanje uporabe mobilne aplikacije za oglaševanje, ki vključuje igrifikacijo in mešano resničnost?

Zadnje raziskovalno vprašanje smo zastavili, saj so v eni od študij [11] ugotovili, da obstaja razlika v razumevanju aplikacij z obogateno resničnostjo med šolarji višjih razredov in študenti. Glavna razlika se je pokazala v načinu razmišljanja, izkušnjah in načinu reševanja problemov.

4.2 Razvoj prototipa aplikacije in predstavitvenih video posnetkov

Prototip smo izdelali s pomočjo programa Adobe XD. Na osnovi izdelanega prototipa smo ustvarili tri videoposnetke, ki so prikazovali vsak po en scenarij primera uporabe aplikacije. Prvi video posnetek je prikazoval osnovno delovanje aplikacije in prvi primer uporabe – nakupovanje v nakupovalnem centru in ujetje željenega popusta. Drugi videoposnetek je prikazoval prejem potisnega sporočila, ki ga uporabnik dobi ob novem popustu za izdelek, ki si ga želi. Tretji videoposnetek pa je prikazoval primer časovno omejenega popusta, ki ga uporabnik mora ujeti v določenem časovnem obdobju ter ga shrani med svoje popuste ter unovči kadar je naslednjič v izbrani poslovalnici. Slika 1 prikazuje zaslonski posnetek iz predstavitvenega video posnetka omenjenega primera.



Slika 1: Slika zaslona predstavitvenega video posnetka.

4.3 Vzorčenje in udeleženci raziskave

Pogoji za vključitev udeležencev v raziskavo so bili naslednji:

- osebe so starejše od 18. leta,
- osebe, ki (vsaj enkrat na mesec) nakupujejo v vsaj eni trgovski verigi,
- osebe, ki imajo v lasti pametni mobilni telefon,

Za vključitev potencialnih udeležencev v raziskavo smo uporabili priložnostno vzorčenje. V raziskavi je sodelovalo 33 udeležencev.

4.4 Merski instrument

Merski instrument, ki smo ga uporabili, je bil spletni anketni vprašalnik. Sestavljen je bil iz treh delov. Prvi del je bil splošnejši in je zajemal vprašanja o sami aplikaciji ter zajemal po eno vprašanje, ki se je navezovalo na eno izmed raziskovalnih vprašanj. Pri RV3 smo navezujoče se vprašanje povezali z demografskim vprašanjem o starosti uporabnika. Drugi del vprašalnika je meril uporabniško izkušnjo – uporabili smo User Experience Questionnaire (UEQ) [12]. Tretji del vprašalnika je meril uporabnost uporabniškega vmesnika – uporabili smo System Usability Scale (SUS) [13].

4.5 Postopek raziskave

Pri načrtovanju in izvedbi raziskave smo sledili Evropskemu kodeksu ravnanja za ohranjanje raziskovalne poštenosti, s čimer smo se zavezali načelu spoštovanja udeležencev raziskave [14]. Prav tako smo upoštevali načela Kodeksa etike in integritete za raziskovalce na Univerzi v Mariboru (Univerza v Mariboru, 2014 - 2020), kodeksa Ameriškega združenja psihologov in kodeksa združenja spletnih raziskovalcev. Spoštovali pa smo tudi Zakon o varstvu osebnih podatkov [15].

Testiranje prototipa je zaradi omejitev osebnih stikov v času pandemije COVID-19 potekalo na daljavo. Udeležencem raziskave smo poslali elektronsko pošto z navodili za izvedbo testiranja. Udeleženci so si v vnaprej določenem zaporedju ogledali tri videoposnetke in na koncu izpolnili tri spletne vprašalnike.

4.6 Statistična obdelava podatkov

Za analizo zbranih podatkov o udeležencih raziskave smo uporabili opisno statistiko, med tem ko smo za analizo podatkov, s pomočjo katerih smo želeli odgovoriti na raziskovalna vprašanja, uporabili tako opisno, kot tudi inferenčno statistiko. Natančneje, odgovore na prvi dve raziskovalni vprašanji smo iskali z opisno statistiko, odgovor na zadnje raziskovalno vprašanje pa z neparametričnim statističnim testom Kruskal-Wallis H Testom. Podatke smo analizirali s programom IBM SPSS Statistics.

4.7 Rezultati

Prvo raziskovalno vprašanje je spraševalo, kakšna je zaznana stopnja vidljivosti oglasa s hkratno vpeljavo igrifikacije in mešane resničnosti v mobilno aplikacijo za oglaševanje. Rezultati deskriptivne statistike so pokazali, da 51,5 % vseh udeležencev meni, da bi oglas nekoliko izstopal, 27,3 % udeležencev pa meni, da bi oglas zelo izstopal. Več kot polovica udeležencev raziskave tako meni, da bi oglas, ki je pripravljen na način kot so ga videli v videoposnetkih, nekoliko izstopal od ostalih načinov oglaševanja.

Drugo raziskovalno vprašanje je spraševalo, kakšna je zaznana stopnja odločitve za nakup izdelka ob hkratni vpeljavi igrifikacije in mešane resničnosti v mobilno aplikacijo za oglaševanje. Udeležence smo spraševali, kako ocenjujejo, da bi jih prikazana aplikacija motivirala k nakupu določenega oglaševanega izdelka [3]. Udeleženci raziskave so lahko izbirali med petimi različnimi odgovori (1 – uporaba bi me zelo motivirala k nakupu, 5 – uporaba me nikakor ne bi motivirala). 57,6 % udeležencev meni, da bi jih uporaba aplikacije nekoliko motivirala k nakupu, 30,3 % pa jih pravi, da jih uporaba ne bi niti bolj, niti manj motivirala [3]. Več kot polovica udeležencev raziskave tako meni, da bi jih uporaba aplikacije nekoliko motivirala k nakupu.

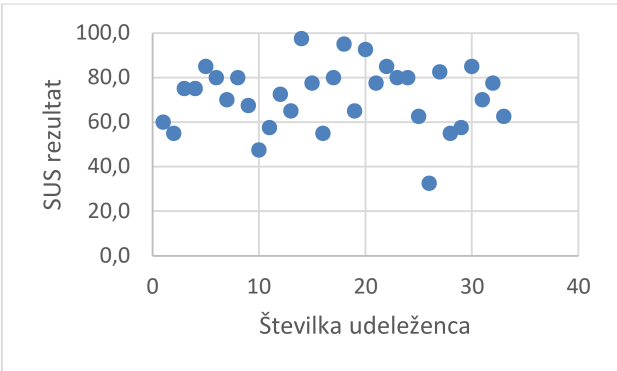
Tretje raziskovalno vprašanje je spraševalo, ali starost uporabnika vpliva na razumevanje uporabe mobilne aplikacije za oglaševanje, ki vključuje igrifikacijo in mešano resničnost. Uporabnike smo razvrstili v štiri starostne skupine: 1 – od 18 do 29 let (19 uporabnikov), 2 – od 30 do 49 let (8 uporabnikov), 3 – od 50 do 64 let (6 uporabnikov), 4 - več kot 65 let (1 uporabnik). Glede na zastopanost v vsaki starostni skupini, smo v analizo vključili prve tri starostne skupine. Rezultati Kruskal-Wallis H testa so pokazali statistično neznačilen rezultat, $p > .05$. S tem lahko sklepamo, da starost uporabnika ne vpliva na razumevanje aplikacije, ki vsebuje igrifikacijo in mešano resničnost [3].

Ocenjevali smo tudi uporabniško izkušnjo razvitega prototipa z UEQ vprašalnikom, kjer so udeleženci ocenjevali svojo uporabniško izkušnjo s pomočjo 26 nasprotujočih si lastnosti. Tabela 1 prikazuje rezultate UEQ lestvice. Razpon lestvice je med -3 in 3, kar označuje izjemno dobro aplikacijo.

Tabela 1: Rezultati UEQ lestvice

Lastnosti	Povprečje	Varianca
Atraktivnost	1,482	0,93
Preglednost	1,508	1,18
Učinkovitost	1,076	0,78
Vodljivost	0,765	0,65
Stimulativnost	0,886	0,84
Originalnost	1,326	0,86

Kot zadnje smo ocenili še uporabnost uporabniškega vmesnika z metodo SUS. Faktor SUS se prikaže na lestvici od 0 do 100. V našem primeru smo izračunali kot povprečno SUS oceno vrednost 71,52. Udeleženci so uporabniški vmesnik prototipa tako ocenili kot dobrega. Slika 2 prikazuje rezultate SUS ocenjevanja.



Slika 2: Rezultati SUS ocenjevanja.

5 DISKUSIJA IN ZAKLJUČEK

Izsledki pričujoče študije se ujemajo z ugotovitvami preteklih raziskav. Tako na primer naši rezultati podpirajo rezultate pretekle študije [8], kjer avtorji ugotavljajo, da je takšen način promocije zanimiv za uporabnika. Prav tako naši rezultati podpirajo rezultate drugih študij [9][10]. V omenjenih študijah namreč ugotavljajo, da se z uporabo mešane ali obogatene resničnosti dviga stopnja zanimanja za nakup iz strani uporabnika.

Omejitve pričujoče raziskovalne študije so v izvedbi testiranja prototipa. Le-ta namreč ni bil testiran na eni lokaciji z več udeleženci.

ACKNOWLEDGMENTS / ZAHVALA

Zahvaljujemo se vsem udeležencem raziskave, brez katerih raziskave ne bi bilo mogoče izvesti.

REFERENCES

[1] Javornik A. Classifications of Augmented Reality Uses in Marketing. IEEE International Symposium on Mixed and Augmented Reality 2014 Media,

Art, Social Science, Humanities and Design Proceedings. 10 - 12 September 2014, Munich, Germany, (2014).

[2] Deterding S., Sicart M., Nacke L., O'Hara K., Dixon D. Gamification. Using game-design elements in non-gaming contexts. International Conference on Human Factors in Computing Systems, CHI 2011, Extended Abstracts Volume, Vancouver, BC, Canada, May 7-12, 2011. (2011), 2425–2428.

[3] Zorko, M. (2020). Združevanje igrifikacije in mešane resničnosti v sodobnem oglaševanju. Fakulteta za elektrotehniko, računalništvo in informatiko. Univerza v Mariboru.

[4] Sailer, M., Hense, J.U., Mayr, S.K., Mandl, H. How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. Comput. Hum. Behav. 69, (2017),371–380.

[5] Moser, T., Hohlagschwandtner, M., Kormann-Hainzl, G., P olzlbauer, S., Wolfartsberger, J. Mixed Reality Applications in Industry: Challenges and Research Areas. SWQD 2019, LNBIP 338, (2019), 95–105.

[6] Nobre, H., & Ferreira, A. Gamification as a platform for brand co-creation experiences. Journal of Brand Management, 24(4), (2017), 349–361.

[7] Teotónio, N., & Reis, J. L. The Gamification Systems Application Elements in the Marketing Perspective. Trends and Advances in Information Systems and Technologies, (2018), 77–87.

[8] Järvinen, S., Peltola, J., & Kemppi, P. Sensor Ball Raffle – Gamification of Billboard Advertising: How to Engage the Audience? Lecture Notes in Computer Science, (2018), 164–174.

[9] Whiskard, H., Jones, D., Voller, S., Snider, C., Gopsill, J., Hicks, B. Mixed Reality Tools as an Enabler for Improving Operation and Maintenance in Small and Medium Enterprises. IFIP Advances in Information and Communication Technology, (2018), 3–14.

[10] Gallardo, C., Rodríguez, S. P., Chango, I. E., Quevedo, W. X., Santana, J., Acosta, A. G., Andaluz, V. H. Augmented Reality as a New Marketing Strategy. Augmented Reality, Virtual Reality, and Computer Graphics, (2018), 351–362.

[11] Klautke H., Bell J., Freer D., Cheng C., Cain W. Bridging the Gulfs: Modifying an EducationalAugmented Reality App to Account for TargetUsers' Age Differences. Springer Nature 2018, (2018), 185-195.

[12] Hinderks, A., Schrepp, M., & Thomaschewski, J. (2018). UEQ-User Experience Questionnaire. Retrieved September, 12, 2018.

[13] Blažica, B., & Lewis, J. R. A Slovene Translation of the System Usability Scale: The SUS-SI. International Journal of Human-Computer Interaction, 31(2), (2015), 112–117.

[14] Evropska znanstvena fundacija. 2011. Evropski kodeks ravnanja za ohranjanje raziskovalne poštenosti. Dostopno na: https://www.arrs.si/sl/analize/publ/inc/Evropski_kodeks_raziskovalne_postenosti.pdf

[15] ZVOP-1. Uradni list RS, št. 94/2007.

StreetGamez: detection of feet movements on the projected gaming surface on the floor

Peter Škrlj
peter.skrlj@student.upr.si
Univerza na Primorskem, UP FAMNIT
Koper, Slovenija

Mark Lochrie
mlochrie@uclan.ac.uk
Media Innovation Studio, University of Central Lancashire
Preston, UK

Matjaž Kljun
matjaz.kljun@famnit.upr.si
Univerza na Primorskem, UP FAMNIT
Koper, Slovenija
Fakulteta za Informacijske Študije
Novo mesto, Slovenija

Klen Čopič Pucihar
klen.copic@famnit.upr.si
Univerza na Primorskem, UP FAMNIT
Koper, Slovenija
Fakulteta za Informacijske Študije
Novo mesto, Slovenija

ABSTRACT

We implemented a software solution for a video game platform that is capable of detecting movement of players' feet on the floor. The solution is a part of a wider project of using a drone as a platform that could project the game board on the floor as well as track movements and scores of different players. The whole system is composed of three parts: a drone, a mini projector, a depth camera and a computational device for running the software. For the latter two we used Google Tango to run spatial recognition, detect 3D shapes and obtain the device's orientation in space. The system was implemented to the point where it can detect the player's feet, transform the detected feet to a gaming surface and correct the projection distortion.

KEYWORDS

exergaming, human-drone interaction, drones, pervasive computing

1 INTRODUCTION

Exercise games or exergames can be divided into three categories: location based games (e.g. [5]), games with motion tracking (e.g. [6]) and projection based games (e.g. [3]). In [4] we proposed a new gaming concept that combines projection based games with drones and user tracking creating a novel gaming platform that is (i) independent of location and (ii) offers a new gaming abilities that can facilitate various types of novel street and chalk games. In this paper we present the software solution for a proposed game platform that is capable of detecting movement of players' feet on the floor.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia

© 2020 Copyright held by the owner/author(s).

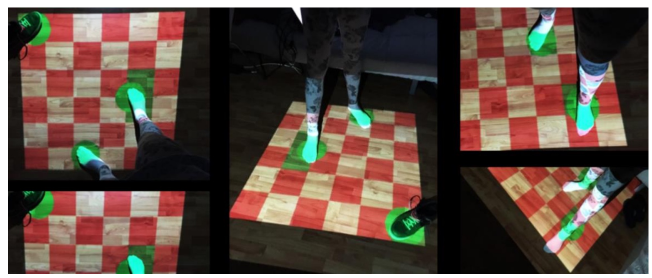


Figure 1: Complete detection system with projection.

2 SYSTEM DESIGN

The minimal set of functional requirements was: (i) Track player's feet on a projected grid where each grid unit measures 30x30 cm to support games such as "whack the mall". A particular unit activates when player steps on it. (ii) Provide quick feedback whilst correctly detecting fast movements, which is vital for an exregaming platform. (iii) The projected surface should be always mapped as a rectangle. To avoid accidents, the drone should hover on the side of the projected surface, which would in this case be distorted and should be corrected. (iv) The platform should support multiple players to increase motivation – an important element of exergaming.

We decided to use the Google Tango device, which is capable of detecting player's movement and control the projection. A variety of other devices could be used to achieve the same. However, at the time when the implementation begun (2016), this was one of the rare devices with such functionalities and light enough for drone carrying. For implementation and testing of the software solution we planned to use the device in a static environment, placed 2m high of the floor facing the playing field under an angle of 45-70 degrees.

Google Tango integrates three main functionalities. Motion tracking of the device by using visual features of its surroundings in combination with the accelerometer and gyroscope. Area learning by recording the visual features and the measuring the distances. Depth perception by scanning and building a point cloud image of the room. From this point cloud, a room meshes can be made and used as 3D models for further processing. This feature is of particular interest to us, as we planed to utilise the depth cloud in order to detect players movement over the ground plane.

To appeal to a wider community of game developers, we decided to use Unity together with Google Tango's SDK to obtain the callback calls and events from the C library used for processing signals in the Tango device itself.

For projection we used the 200 lumens ASUS S3 connected to Google Tango via mini HDMI port. S3 has a wide projection angle capable to project a large playing area from relatively short distance. The image projected has a trapezoidal distortion called keystone distortion caused by the projector projecting at an angle to the projection surface.

3 PROTOTYPE IMPLEMENTATION

The software is built of four (4) components: (i) *floor plane detection* — detecting the ground plane and initialisation; (ii) *point cloud processing and player detection* — searching for players feet position using information from depth camera; (iii) *RGB optimization* — player identification and optimisation of tracking performance; and (iv) *rendering — projector alignment correction* — removing perspective distortions from the projection.

Floor plane detection

There are three common methods for generating depth information: Stereo method using two cameras, Time of Flight casting rays into the space and timing the bounces, and Structured Light. Tango uses the latter using IR projector, which beams a grid pattern of dots where each sample group of the dots is uniquely identified. This way the IR projector and IR camera are able to determine the exact position of the detected point group.

The first step of tracking players is to estimate where ground plane lies. This is done by *floor plane detection* algorithm. After obtaining point cloud data, we start by mapping points into buckets where the Y axis is kept in small deviation groups. At each new point cloud frame the points are added into group and once the threshold is reached, the algorithm marks that Y coordinate as a ground plane. Since the Tango device can localise itself in the space, the ground plane needs to be detected only once at the initialization stage.

Point cloud processing

Once we know the position of the ground plane, we move to *Point cloud processing*, which starts by obtaining point cloud data. Then a simple min max filter on the Y axis can be applied to isolate 3D points that are likely to be feet. We set the filtering threshold to 20 cm distance from the ground plane. Points that fall out of this threshold are discarded. The results of the filtering can be seen in Figure 2.

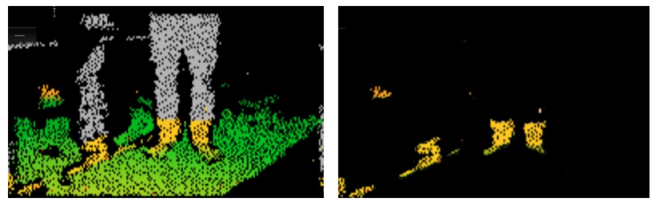


Figure 2: Left — green coloured points are on the floor level, orange points are objects that are within the 20 cm threshold; Right — filtered image after min max filtering.

After filtering, the remaining points are grouped into spherical geometric shapes. This is done by processing every point and trying to fit it into a nearby sphere. The radius of the sphere was manually set to the default diameter of 30 cm. To simplify the grouping process, we ignore the Y coordinate of the feedback mapping placing all feedback points to a single plane. Using a 2D image, we can generate distinct groups by using a simple grouping algorithm. Our algorithm starts the grouping process by randomly selecting a feedback point. Then we check if there is any group defined within the threshold proximity of this point. If not, we create a new point group, set its rank to 1 and set the location of the group to this point. If the point is found in the diameter of an existing group, it is added to the nearest one. The group position is then updated by weighted average as such:

$$GroupPos = \frac{N_i tems \times GroupPos + FeedbackPosition}{N_i tems + 1}$$

After processing all the points, the transformation from 2D group coordinates back to 3D coordinates occurs by averaging the Y coordinate of group points. At this stage we remove groups which consist of insufficient number of detected points. This value can be changed though game engine configuration. The result of this step is an averaged group of strong feedbacks (Figure 3).

RGB optimization

Since the Tango depth camera has a relatively low refresh rate of 10Hz, we planned a fine grain tracking by analysing captured images from RGB camera. To obtain data from the camera the SDK callbacks varied across different versions of

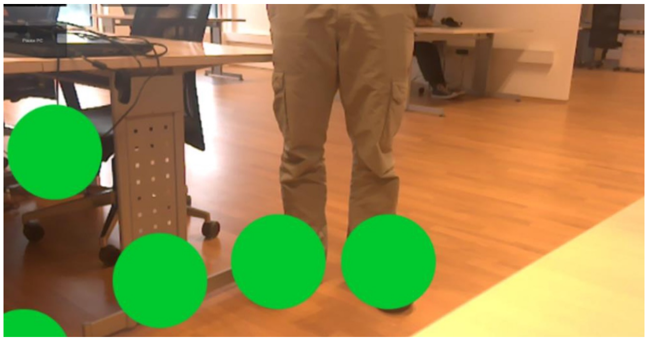
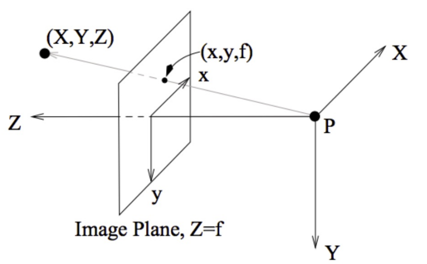


Figure 3: Groups are being rendered back to the scene in form of spheres that cover a certain area in the virtual world.

Tango Core. In Ikariotikos (Version 1.54, June 2017), an event needs to be registered that signals when a new camera image has been rendered to the buffer and is available for reading. Unfortunately, we were not able to obtain the RGB stream whilst depth camera was in operation. The reason for this is still not fully understood and the lack of documentation made it impossible to find the solution within the timeframe of this project. Nevertheless, we present the intended approach for optimising player tracking using color detection.



$$\begin{pmatrix} x \\ y \\ f \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

Figure 4: Pinhole camera model showing how 3D pint is transformed to 2D image.

We planned to perform the colour based tracking within the regions detected by the depth camera whilst waiting for its next frame. The first step was to perform a transformation of detected 3D points from the depth cloud to the screen coordinate system in the step we call "Transformation of points to view port frame". This would allow us to create a mask with regions of interest. Such transformation can be done using a pinhole camera model (Figure 4) [2]. The projection of 3D point cloud to the screen can be therefore formulated as $[xz]_pixel = K \times [XZY]_3DPoint$ where (i) x, y is location of point in image coordinate system, (ii) X, Y, Z is location of points in world coordinate system in which the data is provided, and (iii) K is a matrix of intrinsic camera parameters $\begin{pmatrix} F_x & 0 & F_x \\ 0 & F_z & F_y \\ 0 & 0 & 1 \end{pmatrix}$. To obtain the pixel coordinates within the view port frame, we need to multiply the 3D point with intrinsic camera parameters. These are defined through camera calibration process done by the developer of the device resulting in the coordinate system of the depth camera almost perfectly aligned with the device screen. As such we can ignore extrinsic camera parameters.

After we receive 2D points, we clip the point groups into detection masks for the next step we call "Use the circles as masks to fine track Colour image". In practice we map a 3D vector to camera view port by using Unity WorldToViewPortPoint method call. To proceed, we need to map colour image to the mask by scaling the 2D point so that it corresponds with the captured image. We are then able to cut the detection area from the colour image. To enable adjusting the performance of the detection, the size of the detection square is possible to be manipulated via GUI.

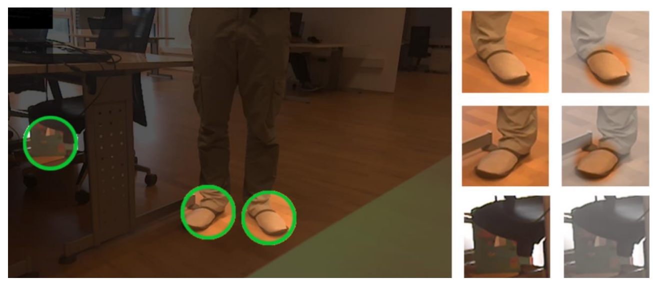


Figure 5: Concept sketch of detecting player feets in RGB image with higher refresh rate.

Using the mask the segments of detection image are cropped out (Figure 5) and the colour group detection is ran over using OpenCV contours finding method, allowing us to filter the colour groups and detect centre and radius [1]. After detecting 2D point groups, we apply Unity methods to transform 2D location on the image to the world coordinates. Since our 3D point detection is also detecting the playing plane, it is possible to calculate the correct point of contact. Mapping the detected centres of the feedback back to the detected floor is an easy task. The points are transformed with an inverse of the mapping of 3D point to the 2D screen space. We simply raycast the screen coordinate of the point to the floor and obtain the group position.

Rendering

Projector alignment correction. Projecting an image to a non perpendicular surface in respect to the light source will produce a distorted image commonly called a keystone effect. This distortion can be approximated by $cos(\varepsilon - \alpha/2)/cos(\varepsilon + \alpha/2)$, where ε is the angle of the surface being projected on, and α is the width of the focus. Because the projector is mounted in the same space as the Tango device, and the device is spatially aware of its orientation in respect to the ground plane, we can calculate the required adjustments to the projected image and project the proper square to the floor. Two rotations that cause the keystone effect are rotation Rz and Rx. Since the depth camera will be pointed at the players, the drone will need to be equipped with a gimbal maintaining the rotation Rx. To successfully correct the distortion caused by Rz we need to know the parameters of the projector's field of view (FOV), lense parameters and Rz in relation to the ground surface.

We created a virtual scene with a single texture we call a Render Texture and a virtual camera to which we assign FOV and aspect ratio that matches our projector. We place the virtual camera at a fixed distance and rotate the plane around the z-axis in the opposite direction of tilt detected by the Tango tracking system. In this way we render graphics where perspective distortions from rotation around the z-axis are removed as seen in Figure 6.

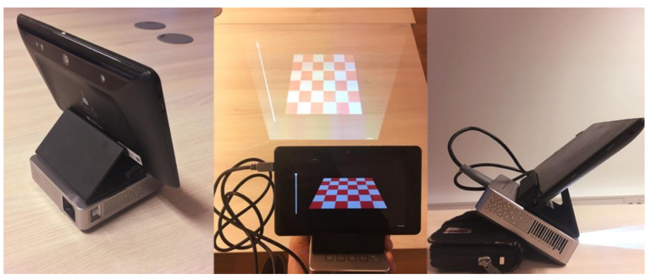


Figure 6: Example of perspective mapping of square onto a flat surface. The internal camera mimics projectors field of view and inverts the projection angle.

This solution only corrects for one rotation, but as Tango device is capable of 6 DOF camera pose tracking, the rotation around x could be accounted for. We could also use tracking information to fix the playing field onto a position in the real world. The playing field would thus stay at the same place regardless of the position and orientation of the drone. A more advanced solution would be to use inverse transformation using game shaders or other transformations possible in Unity. A possible approach would be to apply the correct inverse trapezoid transformation to the image received from Unity.

Mapping feedbacks to a 6x6 playing plane. In the initialization step, a ray is casted from the centre of the camera to the detected floor plane. The intersection of the ray and the plane represents the centre of the detection matrix. Its centre point is used for syncing the display grid with the detection grid. The latter is defined in the engine with default values of 6 columns and 6 rows. This setting can be additionally adjusted to allow more precise feedbacks. However, this may cause performance issues. After the grid is initialised, its fields are updated according to the sphere positions that are being calculated by the point cloud detection algorithm. When a square is overlapped with the sphere it becomes active.

4 PROTOTYPE GAME

Implementing the above, a simple game was created. The detection runs at 10-15 FPS with some lag spikes that additionally occur because of point cloud detection instability. Once the system is initialised (the ground plane is recognised) it starts tracking feet. Where these are tracked green spheres are rendered and segments of the checker box pattern that intersect with the spares are coloured in green (see Figure 1). The playing area in the figure is of size 1,7x1,7 m, meaning that the projected squares were approximately 27x27 cm. The area could be increased by putting the projector and Tango further away. Because the playing field was relatively small, only 2 players could be on it at the same time. The player tracking would fail if there would be more players because of the excessive density of the detected points.

5 CONCLUSIONS

It is important to note that the system is currently limited to projections on horizontal planar surfaces. The optimisation utilising colour tracking of players feet needs to be implemented. Thus, in order to support multiplayer games a unique footwear colour is required for each player. Despite these limitations and the fact that the Tango platform has been deprecated and integrated into Google ARCore, the concepts presented can be utilised for a solution using another platform. More information on the system is available in [7].

REFERENCES

[1] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. 2010. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 33, 5 (2010), 898–916.

[2] A Kaehler and G Bradsk. 2013. Computer Vision in C++ with the OpenCV Library. *O'Relly* (2013).

[3] Raine Kajastila, Leo Holsti, and Perttu Hämäläinen. 2016. The augmented climbing wall: High-exertion proximity interaction on a wall-sized interactive surface. In *CHI '16*. 758–769.

[4] Matjaž Kljun, Klen Čopič Pucihar, Mark Lochrie, and Paul Egglestone. 2015. Streetgamez: A moving projector platform for projected street games. In *CHI in Play '17*. 589–594.

[5] Kate Lund, Paul Coulton, and Andrew Wilson. 2011. Free All Monsters! a context-aware location based game. In *MobileHCI '13*. 675–678.

[6] Emily CS Murphy, Linda Carson, William Neal, Christine Baylis, David Donley, and Rachel Yeater. 2009. Effects of an exercise intervention using Dance Dance Revolution on endothelial function and other risk factors in overweight children. *Internat. Jour. of Pediatric Obesity* 4, 4 (2009), 205–214.

[7] Peter Škrlj. 2017. *StreetGamez, a moving projector platform for games.* Master's thesis. University of Primorska, 6000 Koper.

Anamorfična projekcija na poljubno neravno površino

Anamorphic projection on an arbitrary uneven surface

Rok Cej
Franc Solina
rokcej1997@gmail.com
franc.solina@fri.uni-lj.si
Laboratorij za računalniški vid
Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Večna pot 113
1000 Ljubljana, Slovenia

POVZETEK

Razvili smo metodo, ki omogoča anamorfično projekcijo na neravno, razbrazdano površino. Sliko, ki jo projeciramo v tem primeru, ni dovolj le v celoti perspektivno deformirati. Neravna površina je namreč sestavljena iz velikega števila majhnih ploskev različnih orientacij in za vsako od teh ploskev bi morali izračunati ustrezno perspektivno deformacijo. To najlažje storimo tako, da za vsak slikovni element projecirane slike izračunamo ustrezno deformacijo. To pa zahteva, da imamo 3D model površine, na katero se slika projecira, kar pridobimo s pomočjo senzorja "Kinect".

KLJUČNE BESEDE

Anamorfoza, Kinect, globinski senzor, optična iluzija

ABSTRACT

This report describes the creation of a distorted image or video that looks perfect when projected onto a given uneven surface and viewed from a predetermined angle. It utilizes the depth sensor Kinect and a projector. The program is written in C++ and it starts off by recreating the projection surface in 3D. It then uses the surface model to create an anamorphic projection. If the Kinect and the projector are properly aligned, the projected image or video creates an anamorphic illusion in real life.

KEYWORDS

Anamorphosis, Kinect, depth sensor, optical illusion

1 UVOD

Ljudje lahko dokaj zanesljivo interpretiramo slike, ki jih ne gledamo frontalno, ampak pod določenim kotom, saj zna naš zaznavni sistem podzavestno razstaviti informacijo na vsebino slike in na njeno perspektivno deformacijo. Še posebej dobro ta princip deluje, če lahko zanesljivo zaznamo, kako je slikovna ploskev orientirana v prostoru. Pri tem igra pomembno vlogo tudi koherenca med premikanjem opazovalca in perspektivno deformacijo. Majhen premik opazovalca povzroči le majhno spremembo perspektivne deformacije. Pri anamorfičnih slikah pa ta koherenca ne obstaja. Anamorfična podoba se tipično razkrije le iz točno določene smeri opazovalčevega pogleda. Odvisno od vrste anamorfoze, je ta smer pogleda lahko bolj ali manj natančno določena.

1.1 Vrste anamorfoz

Anamorfozo so odkrili v času renesanse, ko so umetniki in znanstveniki odkrivali zakone perspektive [2, 3]. Prva vrsta anamorfoze, ki so jo uporabljali, je bila **perspektivna anamorfoza**. Perspektivno deformirana podoba je naslikana na ravno ploskev. Da bi se ta anamorfična podoba razkrila, jo je potrebno pogledati z določenega zornega kota, običajno je to dokaj oster kot glede na ravnino, ki nosi deformirano podobo (Slika 1).

Katoprične ali zrcalne anamorfoze za razkritje prave podobe potrebujejo ogledalo, običajno cilindrične ali konične oblike. Če tako ogledalo postavimo na pravo mesto, se deformirana podoba razkrije kot odsev v ogledalu (Slika 2).

Med anamorfične upodobitve štejemo tudi **iluzionistično slikarstvo**, kjer lahko na predvidenem mestu opazovanja pridobimo izrazit občutek prostorske dimenzije. V umetnostni zgodovini so znane predvsem poslikave stropov, kjer se nam dozdeva, da se prostor odpira proti nebu (Slika 3), danes pa podoben prostorski učinek uporabljajo potujoči umetniki, ki s kredo rišejo podobe na ulicah (Slika 4).

Sodobni umetniki, kot je npr. švicarski slikar Felice Varini [11], anamorfozo uporabljajo pri poslikavi notranjih prostorov ali celih urbanih scen tako, da se z določenega zornega kota razkrije nek pravilen geometrijski vzorec, kot da bi lebdel v prostoru (Slika 5). Anamorfični princip se uporablja tudi pri slikanju prometnih označb na cestišča, da bi bila bolj jasno berljiva in razločna pod ostrim kotom opazovanja, kot ga imajo vozniki in drugi udeleženci v prometu. Tudi razni reklamni napisi, ki jih pravilno vidimo v zrcalih ali pod določenim kotom opazovanja sodijo v kategorijo anamorfičnih poslikav.

S pojavom multimedijske tehnologije se je pojavila možnost, da za prikaz anamorfičnih upodobitev uporabimo video projekcijo. Na primer, reklamne napise je možno perspektivno deformirati, tako da njihova projekcija iz notranjosti trgovin na pločnik pred trgovino ni deformirana in je zato lažje berljiva.

V Laboratoriju za računalniški vid smo celo razvili princip **dinamične anamorfoze**, ki perspektivno deformacijo projecirane slike stalno stalno prilagajajo poziciji opazovalca, tako da je z opazovalčevega zornega kota slika stalno izgleda nedeformirana oziroma tako, kot če bi jo gledali frontalno [8].

2 MOTIVACIJA

Če uporabljamo video projektor, je projecirana slika brez vsakršne perspektivne deformacije le, če jo gledamo natanko iz točke projeciranja. Ker ima projektor svoje fizične dimenzije, to v praksi seveda ni možno in zato je projecirana slika, ki jo gledamo vedno

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

Slika 1: Ena od najbolj znanih slik iz zgodovine umetnosti, ki upodablja perspektivno anamorfozo, sta *Ambasadorja* Hansa Holbeina iz leta 1533. Lobanja, ki se v frontalnem pogledu (levo) vidi kot eliptičen madež na sredini slike spodaj, pa se v pogledu od desno zgoraj (v sredini), razkrije kot lobanja (desno). Umetniki so tako ekstremno popačenje običajno uporabili, da bi skrili določene kontroverzne elemente na sliki (vir: Wikimedia Commons).

Slika 2: Zrcalna anamorfoza: popačena 3D skulptura se v odsevu cilindričnega zrcala razkrije kot žaba (avtor: Jonty Hurwitz, vir: Wikimedia Commons).

Slika 3: Primer iluzionistične poslikave stropa je v celjski Stari grofiji, ki ga umeščajo na prehod iz renesanse v zgodnji barok (vir: Wikimedia Commons).

nekoliko deformirana. Kot smo že v uvodu razložili, to običajno ni problem, saj človeška zaznava z lahkoto loči med informacijo na sliki in zmerno perspektivno deformacijo te iste slike. Če pa je kot med osjo projekcije in smerjo našega pogleda zelo velik, pa že lahko nastopijo težave pri interpretaciji slike. Pri anamorfozi pa na ta način pravzaprav želimo skriti pravi pomen slike ali vsaj dela slike. Še večji problem pri interpretaciji slike nastane, če projekcijska površina ni ravna. Zato je naš raziskovalni motiv naslednji – *ali lahko projecirano sliko vnaprej deformiramo tako, da bo izgledala nedeformirano iz vnaprej določenega zornega kota, neglede na to, kakšna je površina, na katero projeciramo sliko?* Z drugimi besedami, kako lahko izračunamo inverzno anamorfično deformacijo slike, da bo izgledala pravilno na poljubni neravni površini?

Slika 4: Uporaba perspektivne anamorfoze v uličnem slikarstvu (avtor: Julian Beever, 1990-ta). Na levi se vidi izrazit prostorski učinek, gledano z nasprotne strani, pa se vidi kako popačena je na tlaku dejanska podoba, še posebej izrazito noga kopalke, ki v 3D iluziji sega najdlje iz slikovne ploskve (vir: Wikimedia Commons).

Slika 5: Ploskovna grafika superponirana na razgibano urbano sceno, se v celoti razkrije le s točno določenega zornega kota: Felice Varini, Port de St-Nazaire, Francija, za razstavo “Estuaire 2007” (vir: Wikimedia Commons).

Že pri običajni perspektivni anamorfozi moramo vedeti, kako je slikovna ploskev orientirana v prostoru. Če pa želimo sliko projecirati na poljubno neravno površino, moramo imeti 3D model te površine. Sodobna tehnika ima za odčitavanje 3D oblik številne odgovore. Cenovno ugodna in za naše potrebe je smiselna uporaba senzorja Microsoft Kinect. Kinect smo v našem laboratoriju že uporabili za odčitavanje 3D površine v sorodnem projektu *Svetlobni vodnjak* [9], kjer smo klasični kamniti skulpturi dodali še virtualno dimenzijo v obliki polzečih vodnih kapljic, ki smo jih z video projektorjem projecirali v obliki svetlobnih pik [10].

3 SORODNA DELA

Na prvi pogled je naš cilj najbolj podoben tehnikam, ki s pomočjo video projekcije na 3D predmete (angl. projection mapping [12]) ustvarijo obogateno resničnost in tako omogočijo povsem novo in dodatno dimenzijo dojemanja tudi gibajočih se predmetov, npr. [4]. Vendar se naš problem razlikuje od zgoraj opisanega v dveh bistvenih elementih:

(1) Nam ni potrebno video projekcije poravnati z neko vnaprej določeno 3D obliko oziroma predmetom. Zato kompleksna geometrijska kalibracija med 3D površino, na katero se projecira in katere obliko zajema globinski senzor, ter video projekcijo ni potrebna [5].
(2) V večini sistemov za video obogateno resničnost je smer gledanja uporabnika v grobem poravnana s smerjo video projekcije in zato do potrebe ali pojava perspektivne anamorfoze niti ne pride, čeprav s sledenjem položaja uporabnika nekateri sistemi tudi ustrezno korigirajo pespektivno deformacijo v video projekciji [6].

V komercialnih sistemih za video obogateno resničnost, npr. [7], so tudi integrirani globinski senzorji, vendar ti služijo predvsem avtomatični segmentaciji scene na osnovi oddaljenost od projektorja, da zamudna ročna segmentacija slike ni več potrebna. Zato smo se odločili za razvoj lastnega sistema za anamorfno projekcijo na neravno površino, ki je namenjen opazovanju projekcije iz nekega vnaprej določenega zornega kota.

4 OPREMA

Za anamorfično projekcijo na poljubno neravno površino potrebujemo dve zunanji napravi: Microsoft Kinect in video projektor. Kinect meri razdalje med 0,5m in 4,5m, kar narekuje tudi naš delovni prostor za projekcijo anamorfoze.

Programsko opremo za deformacijo slike smo zaradi hitrosti izvajanja razvili v jeziku C++, čeprav bi po funkcionalnosti bila primerna tudi visokonivojska jezika kot sta Processing in Python. Uporabili smo naslednje knjižnice:

- **OpenGL**: Aplikacijski programski vmesnik (API) za grafiko
  - **GLFW**: kreiranje okolja OpenGL
  - **GLEW**: nalaganje razširitev OpenGL
  - **GLM**: matrične in vektorske aplikacije
- **Kinect SDK**: API za Kinect
- **FFmpeg**: dekodiranje video zapisov
- **stb_image**: branje slikovnih datotek

5 PERSPEKTIVNA ANAMORFOZA NA NERAVNO POVRŠINO

Postopek za inverzijo anamorfične deformacije slike smo razdelil na več korakov.

Pridobivanje globinske slike. Globinske slike, ki jih pridobiva Kinect imajo dimenzijo 512 × 424, slikovne pike pa imajo celoštevilske vrednosti, ki so predstavljene s 16 biti. Vsaka od teh vrednosti predstavlja razdaljo izraženo v milimetrih. Če te vrednosti preslikamo v sivinsko sliko, dobimo globinsko sliko, kjer so v našem primeru svetle točke bolj oddaljene od senzorja. Kjer Kinect ni mogel zajeti globine, so točke črne barve.

Aproksimacija manjkajočih globinskih podatkov. Ker Kinect ne more zajeti globine v vsaki točki bodisi zato, ker je bodisi točka preveč oddaljena, ker se infrardeča svetloba, ki jo Kinect uporablja, odbije od površine ali zaradi šuma. Manjkajoče vrednosti določimo z aproksimacijo na osnovi sosednih točk.

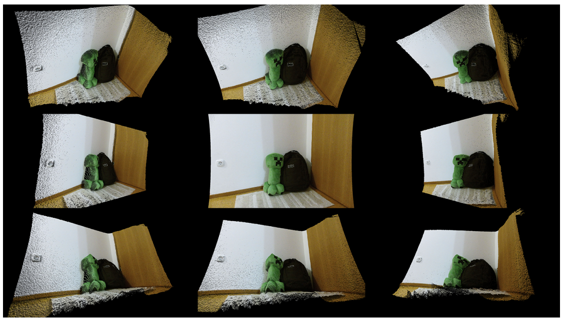
Konverzija globinske slike v oblak 3D točk. Vrednosti posameznih slikovnih točk v globinski sliki spremenimo v koordinate 3D točk z naslednjo enačbo:

$$\vec{position} = depth * \begin{bmatrix} (\frac{2x}{width-1} - 1) * tan(\frac{fov_x}{2}) \\ (\frac{2y}{height-1} - 1) * tan(\frac{fov_y}{2}) \\ 1 \end{bmatrix} \quad (1)$$

kjer je:

$depth$ = globina
x, y = indeks točke v globinski sliki
$width, height$ = resolucija senzorja v hor. in vert. smeri
fov_x, fov_y = zorni kot Kinecta v hor. in vert. smeri v radianih

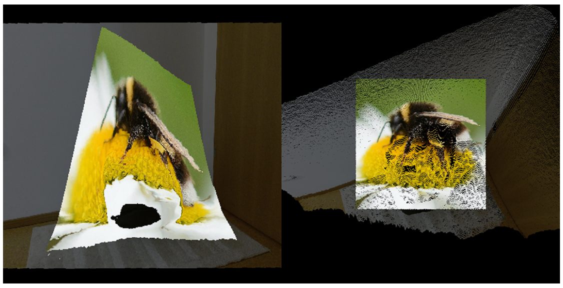
Ker ima Kinect tudi barvno kamero, lahko poveže globinske točke s ustreznimi barvnimi vrednostmi iz barvne kamere. Zato lahko te barve pripišemo tudi 3D točkam. Na sliki 6 je pogled na oblak 3D pobarvanih točk z različnih zornih kotov.



Slika 6: Pogled na oblak točk z različnih zornih kotov.

Virtualna anamorfoza. Najprej bomo izračunali virtualno anamorfozo v virtualnem prostoru, preden to naredimo v realnem prostoru. Najprej predpostavimo, da imamo virtualnega opazovalca, ki gleda v smeri pravokotno na smer projekcijskega snopa. Nato si predstavljajmo, da ta opazovalec projecira sliko na razgibano projekcijsko površino. Ta slika bo za opazovalca izgledala povsem pravilno, toda iz smeri projektorja bo popačena. Za vsako točko v oblaku 3D točk, ki predstavlja projekcijsko površino, izračunamo smer med opazovalcem in to točko in ugotovimo, kje ta premica prebada projecirano sliko. Na ta način določimo korespondenco med vsako točko v oblaku 3D točk in ustreznim pikslom projecirane slike. Ko 3D točkam pripišemo korespondenčno teksturo iz slike, se v oblaku 3D točk pojavi popačena slika, vendar če na oblak pogledamo iz smeri virtualnega opazovalca, dobimo nepopačeno sliko (slika 7).

Prava anamorfoza. Da bi dosegli isti učinek tudi v realnem svetu, moramo sedaj izračunati sliko, ki naj jo projecira projektor, da bi opazovalec videl nepopačeno sliko. Za vsak piksel



Slika 7: Virtualna anamorfoza: pogled iz smeri projektorja (levo) in pogled iz smeri virtualnega opazovalca (desno).

projecirane slike izračunamo smer v katero se ta piksel projecira v 3D prostoru. Zanima nas, kje je presečišče med to smerjo in projekcijsko površino, ki pa je predstavljena kot oblak 3D točk. Dodaten problem povzroča še različna resolucija projecirane slike, ki je veliko višja od resolucije globinskega senzorja (Kinecta), ki definira oblak 3D točk. Zato večina pikslov projecirane slike ni imela direktne korespondenčne 3D točke, ampak smo morali iz štirih najbližjih 3D točk izračunati približek presečišča. Za vsako presečiščno točko smo nato, upoštevaje pozicijo virtualnega opazovalca, lahko povezali piksle projecirane slike z ustreznim pikslom na sliki.

Ker je ta postopek dokaj zamuden, smo uporabili večnitno procesiranje, saj je določanje vrednosti posameznih pikslov v projecirani sliki, neodvisno drug od drugega. Primer tako izračunane projecirane slike je na sliki 8.

Slika 8: Anamorfoza v oblaku 3D točk (levo) in projecirana anamorfično deformirana slika (desno).

Kalibracija. Preden posnamemo 3D model površine za projeciranje ga moramo kalibrirati z video projektorjem. Implementirali smo funkcijo, ki na oblak 3D točk nariše rdeč pravokotnik, ki predstavlja področje, za katerega Kinect pričakuje, da bo nanj projecirana slika. Uporabnik mora nato ročno poravnati pozicijo Kinecta ali video projektorja tako, da se rdeči pravokotnik poravna s projecirano sliko.

6 REZULTATI IN ZAKLJUČEK

Slika 9 prikazuje projekcijo fotografije v horizontalni smeri na nagnjeno razbrazdano kamnito površino in pogled na to projekcijo navpično navzdol, kjer se anamorfoza razkrije — proporci slike so enaki kot na originalni fotografiji. Program na zmogljivem osebnem računalniku teče dovolj hitro, da lahko v realnem času procesiramo tudi video [1].

Zaradi nenatančnosti pri zajemu globinske slike je v anamorfični sliki še nekaj nenatančnosti, kar bi bilo možno preseči z bolj

Slika 9: Levo: originalna slika; Sredina: projecirana slika na nagnjeno, neravno površino; Desno: pogled na projecirano sliko navpično navzdol.

natančnim globinskim senzorjem. Vseeno pa je tak način video projekcije na poljubno neravno površino možno uporabiti za številne aplikacije. Če bi v živo zajemali globinsko sliko, kar Kinect nenazadnje omogoča, bi bilo možno projecirati nedeformirane slike in video tudi na gibajoče se tarče.

ZAHVALA

Raziskovalni program Računalniški vid št. P2-0214 (B) je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

LITERATURA

[1] Rok Cej. *Demonstracija anamorfoze na neravno površino (video).* 2020. URL: http://youtu.be/_eypZlZTRcM (pridobljeno 10. 9. 2020).

[2] Daniel L Collins. "Anamorphosis and the Eccentric Observer: History, Technique, and Current Practice". V: *Leonardo* 25.2 (1992), str. 179–187.

[3] Daniel L Collins. "Anamorphosis and the Eccentric Observer: Inverted Perspective and Construction of the Gaze". V: *Leonardo* 25.1 (1992), str. 72–82.

[4] Creators. *Box.* 2013. URL: https://youtu.be/lX6JcybgDFo (pridobljeno 24. 9. 2020).

[5] Anselm Grundhöfer in Daisuke Iwai. "Recent advances in projection mapping algorithms, hardware and applications". V: *Computer Graphics Forum.* Zv. 37. 2. Wiley Online Library. 2018, str. 653–675.

[6] Brett Jones in sod. "RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-Camera Units". V: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology.* 2014, str. 637–644.

[7] *Lightform, Design Tools for Projection.* 2020. URL: http://lightform.com (pridobljeno 10. 9. 2020).

[8] Robert Ravnik in sod. "Dynamic anamorphosis as a special, computer-generated user interface". V: *Interacting with computers* 26.1 (2014), str. 46–62.

[9] Franc Solina in Blaž Meden. "Light fountain–a virtually enhanced stone sculpture". V: *Digital Creativity* 28.2 (2017), str. 89–102.

[10] Solina, Franc. *Light Fountain 2 - Galaxy.* 2018. URL: http://youtu.be/y6NAiXlNm20 (pridobljeno 10. 9. 2020).

[11] Wikipedia contributors. *Felice Varini — Wikipedia, The Free Encyclopedia.* 2020. URL: http://en.wikipedia.org/w/index.php?title=Felice_Varini&oldid=953793776 (pridobljeno 10. 9. 2020).

[12] Wikipedia contributors. *Projection mapping — Wikipedia, The Free Encyclopedia.* 2020. URL: https://en.wikipedia.org/wiki/Projection_mapping (pridobljeno 10. 9. 2020).

Učinkovita predstavitev slovarskih jezikovnih virov pri govornih tehnologijah

Jerneja Žganec Gros
Alpineon d.o.o.
Ulica Iga Grudna 15
1000 Ljubljana, Slovenija
jerneja.gros@alpineon.si

Žiga Golob
Alpineon d.o.o.
Ulica Iga Grudna 15
1000 Ljubljana, Slovenija
ziga.golob@alpineon.si

Simon Dobrišek
Univerza v Ljubljani, FE
Tržaška cesta 25
1000 Ljubljana, Slovenija
simon.dobrisek@fe.uni-lj.si

POVZETEK

Končni pretvorniki predstavljajo kompakten način za predstavitev slovarjev izgovarjav, ki jih potrebujemo pri sintezi ali prepoznavi govora. V članku je predstavljena nadgradnja končnih pretvornikov, t.i. končni super pretvorniki, s katerimi lahko razširjeni slovar izgovarjav predstavimo z manjšim številom stanj in prehodov kot s pomočjo minimalnega determinističnega končnega pretvornika. Končni super pretvornik ohranja determinističnost, poleg besed iz slovarja lahko dodatno sprejme tudi nekatere druge, neznane besede. Pri tem so lahko oddani izhodni alofonski prepisi za določene neznane besede napačni, vendar se izkaže, da je napaka primerljiva s trenutno najboljšimi metodami za določanje grafemsko-alofonske pretvorbe.

KLJUČNE BESEDE

govorne tehnologije, jezikovni viri, sinteza govora, slovarji izgovarjav

1 Uvod

Govorno podprti uporabniški vmesniki omogočajo uporabniško prijazno interaktivno komunikacijo, še posebej v okolju mobilnih komunikacij. Sodobni koncepti sistemov govorne komunikacije se v praksi prenašajo na majhne prenosne naprave, ki so zasnovane na vgrajenih sistemih (angl. embedded systems), za katere sta značilna omejena procesorska moč ter pomnilniška zmogljivost. Za uspešen razvoj in uporabo govorno podprtih aplikacij na prenosnih napravah je potrebno zagotoviti učinkovite in visoko kakovostne komponente sistema govornega dialoga, to je uspešnost avtomatskega razpoznavanja govora in kakovostno, razumljivo in naravno zveneče sintezo govora.

Implementacija predstavitve leksikalnih jezikovnih virov v celovitih sistemih za prepoznavanje ali sintezo govora na vgrajenih platformah predstavlja netrivialen problem, ki ga še dodatno otežujejo omejitve zaradi uporabljene strojne opreme.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

Pregled znanstvene literature pokaže, da pri izgradnji govornih tehnologij za jezike z velikim številom pregibnih oblik uporaba postopkov, ki so bili razviti za angleški jezik, ni učinkovita (Golob, 2012). Zaradi velikega števila pregibnih oblik besed pri istem številu leksemov se obsežnost prepoznavalnika slovenskega govora v primerjavi s primerljivim prepoznavalnikom angleškega govora vsaj podeseteri. Zato je potrebno posebno pozornost posvetiti prav optimizaciji uporabljenih modelov in njihovi adaptaciji na morfološke posebnosti pregibno bogatih jezikov.

Pomemben del govorno tehnološke aplikacije, kot je denimo sintetizator govora, predstavlja sistem za pretvorbo grafemskega zapisa besed v alofonski prepis. Samodejno določanje alofonskega prepisa v slovenščini temelji na množici kontekstno odvisnih pravil, pri čemer moramo poznati besedni naglas (Gros in Mihelič, 1999). Samodejno določanje besednega naglasa slovenskih besed zaradi nepredvidljivosti naglasnega mesta predstavlja zahtevno nalogo (Golob, 2009), zato je za kvalitetno sintezo slovenskega govora nujna uporaba obsežnih slovarjev izgovarjav.

Slovar izgovarjav predstavlja preslikavo grafemskih zapisov besed v alofonske prepise. Pri pregibno bogatih jezikih, kot je slovenščina, lahko slovarji vsebujejo več milijonov slovarskih vnosov, zaradi česar je lahko njihova uporaba v pomnilniško manj zmogljivih sistemih, kot so npr. vgrajeni sistemi, problematična. V teh primerih je nujna uporaba postopkov, ki omogočajo pomnilniško učinkovito predstavitev slovarjev.

Zato smo želeli poiskati in preizkusiti učinkovite postopke za zmanjševanje odvečnosti pri predstavitvi in računalniškem zapisu jezikovnih virov za pregibno bogate jezikovne skupine, ki bodo omogočali hitro, pomnilniško čim manj zahtevno ter visokokakovostno pretvorbo grafemskega zapisa besed v fonetični prepis in obratno.

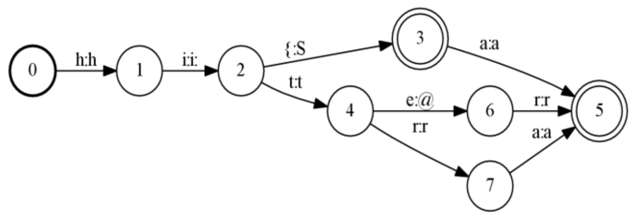
V literaturi je mogoče zaslediti predvsem tri metode, ki omogočajo pomnilniško učinkovito predstavitev slovarjev izgovarjav, in sicer s pomočjo oštevilčenih *končnih avtomatov* (Lucchesi in Kowaltowski, 1993; Daciuk in Piskorski, 2011), dreves predpon (Ristov, 2005) ter *končnih pretvornikov* (odslej kratko *KP*) (Mohri, 1994; Golob at al., 2012). V tem delu bomo predstavili nov način predstavitve s pomočjo *končnih super pretvornikov* (odslej kratko *KSP*), ki predstavljajo nadgradnjo KP. Poleg manjše predstavitve slovarjev v primerjavi s KP, lahko s KSP z visoko točnostjo določimo alofonski prepis tudi nekaterim neznanim besedam oz. besedam, ki niso vsebovane v izvirnem slovarju izgovarjav.

V članku bomo najprej na kratko predstavili KP ter prikazali, kako lahko z njimi predstavimo slovar izgovarjav. Nadalje bomo pokazali, da zastopanost pregibnih oblik v slovarju močno vpliva na velikost KP. Sledi predstavitev KSP, ki predstavljajo nov način predstavitve slovarjev. Rezultate bomo predstavili na jezikovnih virih, ki so bili nadgrajeni v okviru projekta OptiLEX.

2 Končni pretvorniki (KP) ter predstavitev slovarjev izgovarjav

KP sestavljajo stanja ter prehodi med stanji. Vsak prehod ima vhodno in izhodno oznako. Ko se na vhodu KP pojavi določen vhodni niz, se ta nahaja v začetnem stanju. KP nato po vrsti sprejema vhodne simbole. Pri vsakem sprejetju vhodnega simbola odda izhodni niz simbolov, ki ga določa izhodna oznaka pripadajočega prehoda, ter se premakne v naslednje stanje. Če za poljuben vhodni simbol v trenutnem stanju ne obstaja prehod, ki ima vhodno oznako enako temu simbolu, pravimo, da KP vhodnega niza ne sprejema. Če se KP po prejetju vseh simbolov vhodnega niza nahaja v končnem stanju, pravimo, da vhodni niz sprejema, pri tem pa postane oddan izhodni niz veljaven. Omenimo še to, da je lahko vhodna ali/in izhodna oznaka enaka praznemu simbolu oziroma nizu.

KP, ki imajo v poljubnem stanju največ en prehod z določeno vhodno oznako, pravimo deterministični KP. Za takšne KP je hitrost pretvorbe vhodnega niza v izhodni niz zelo hitra in ob primerni izvedbi odvisna samo od dolžine vhodnega niza. Druga prednost determinističnih KP je ta, da obstajajo učinkoviti algoritmi za njihovo minimizacijo. Tako dobimo minimalni KP, ki ima najmanjše število prehodov in stanj med vsemi ekvivalentnimi KP (Mohri, 1997), torej KP, ki za poljuben sprejet vhodni niz oddajo enak izhodni niz.



Slika 1: Primer KP, ki predstavlja slovar izgovarjav za štiri slovenske besede: ***hiš, hiša, hiter*** **in** ***hitra*****. Krogi predstavljajo stanja, puščice pa prehode med stanji. Vsak prehod je označen z vhodno in izhodno oznako, ki sta ločeni z dvopičjem. Začetno stanje je označeno z odebeljenim krogom, končna stanja pa z dvojnim krogom.**

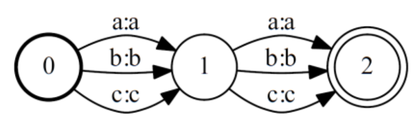
Vseh KP ni mogoče determinizirati, saj imajo deterministični KP manjšo izrazno moč kot nedeterministični (Hellis, 2004). KP, ki predstavlja slovar izgovarjav, lahko vedno determiniziramo, če iz slovarja odstranimo enakopisnice. Slika 1 prikazuje primer *minimiziranega in determiniziranega KP* (odslej kratko *MDKP*), ki predstavlja slovar za štiri slovenske besede.

3 Vpliv velikosti slovarja izgovarjav na velikost končnega pretvornika KP

V tem eksperimentu smo želeli preveriti odvisnost velikosti KP od velikosti slovarja, ki ga želimo predstaviti. Na voljo smo imeli slovar SI-PRON za slovenski jezik, ki vsebuje več kot milijon različnih slovarskih vnosov (Žganec-Gros et al., 2006). Slovar smo razširili z dodatnimi leksikalnimi enotami, ki smo jih razvili v okviru projekta OptiLEX.

Z naključnim izbiranjem slovarskih vnosov smo zgradili več pod-slovarjev različnih velikosti in za vse pod-slovarje zgradili MDKP.

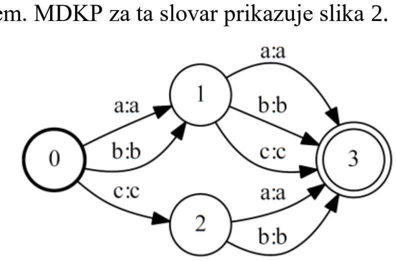
Velikost MDKP, je podobno kot v eksperimentu, izvedenemu na manj obsežnih jezikovnih virih (Golob et al., 2012), dosegla vrh pri 78% do 83% velikosti prvotnega slovarja. To pomeni, da začne velikost MDKP pri določeni velikosti z dodajanjem novih besed oz. slovarskih vnosov iz slovarja upadati. To opažanje je bilo motivacija za razvoj nove vrste končnih pretvornikov, ki jim pravimo končni super pretvorniki., in jih opisujemo v naslednjem razdelku.



Slika 2: MDKP za izmišljen slovar, katerega ključi so sestavljeni iz vseh možnih izborov dveh črk od treh možnih – ***a*****,** b **in** ***c*****. Pri tem so vrednosti enake ključem.**

Da bi si ta pojav lahko lažje predstavljali, poglejmo minimalni primer, ki prikazuje mehanizem tega zmanjšanja velikosti MDKP. Slovarski vnosi so sestavljeni iz para ključ, vrednost. Pri slovarju izgovarjav tako grafemski zapis predstavlja ključ, alofonski prepis pa vrednost.

Kot primer vzemimo vzorčni slovar, katerega ključi so sestavljeni iz vseh možnih izborov dveh črk od treh možnih, npr. črk *a*,*b* in *c*. Na ta način dobimo 9 različnih ključev, in sicer: *aa, ab, ac, ba, bb*… Zaradi enostavnosti naj bodo pripadajoče vrednosti enake ključem. MDKP za ta slovar prikazuje slika 2.



Slika 3: MDKP za enak slovar, kot ga predstavlja MDKP na sliki 2, pri čemer mu manjka slovarski vnos ***cc*** **:** ***cc*****.**

Sedaj iz omenjenega slovarja odstranimo slovarski vnos *cc* : *cc* ter ponovno zgradimo MDKP. Rezultat prikazuje slika 3. Opazimo lahko, da se je pri odstranitvi slovarskega vnosa iz slovarja kompleksnost MDKP povečala, saj je za predstavitev slovarja potrebno eno dodatno stanje ter dva dodatna prehoda. V (Golob et al., 2012) in (Golob et al., 2016) smo podrobneje raziskali vzroke, ki vplivajo na zmanjšanje MDKP pri predstavitvi slovarja pri dodajanju novih slovarskih vnosov.

3.1 Vpliv množičnosti pregibnih oblik na velikost slovarja izgovarjav

Preverili smo vpliv množičnosti pregibnih oblik lem besed iz slovarja na velikost MDKP. Pri tem z množičnostjo pregibnih oblik mislimo na število različnih pregibnih oblik za določeno lemo. Za primer smo vzeli besedo *skopati* ter v slovarju poiskali vse slovarske vnose, katerih grafemski zapisi predstavljajo pregibne oblike leme izbrane besede. Dobili smo 27 različnih slovarskih vnosov, iz katerih smo s pomočjo naključnega izbiranja vnosov tvorili še štiri različno velike pod-slovarje. Za vsak pod-slovar smo zgradili MDKP. Iz rezultatov je razvidno, da hitrost naraščanja velikosti MDKP z večanjem slovarja rahlo pada, vendar pa ni opaziti obrata trenda povečevanja MDKP (Golob et al., 2012).

3.2 Vpliv množičnosti pregibnih oblik na velikost slovarja izgovarjav

Poglejmo sedaj, kako na velikost MDKP vpliva zastopanost pregibnih oblik v slovarju, sestavljenem iz večih besed, ki se podobno pregibajo. Iz slovarja SI-PRON smo izbrali 28 grafemskih zapisov besed, katerih pregibne oblike imajo 9 različnih končnic ter pripadajo štirim različnim lemam - *potop, osmod, zasp, natoč*. Izbrane leme ter pripadajoče končnice so prikazane v tabeli 1. Lema *zasp* pri tem predstavlja izjemo, ki se pregiba nekoliko drugače kot ostale tri.

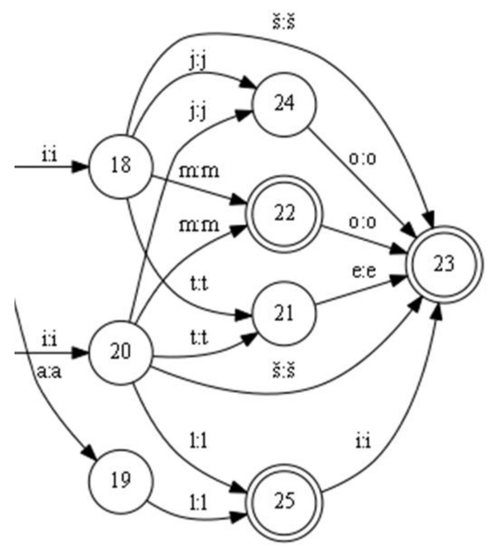
Tabela 1: Tabela prikazuje postopek za tvorjenje vseh besed, ki so vsebovane v slovarju. V levem stolpcu so navedene leme besed, v desnem pa možne končnice.

MOŽNE LEME	MOŽNE KONČNICE
potop, osmod, zasp, natoč	iš,im,imo,ite,ijo
potop, osmod,natoč	i+l,i+li
zasp	a+l, a+li

Iz teh besed smo nato tvorili slovar, pri čemer smo zaradi enostavnosti vrednosti ključev izenačili s ključi. Nato smo z naključnim izbiranjem iz tega slovarja tvorili še štiri različno velike pod-slovarje. Za vse tako zgrajene slovarje smo zgradili MDKP.

Velikost MDKP, ki predstavlja vseh 28 vnosov slovarja, manjša od MDKP, ki predstavljata slovarja s 23 in 17 vnosi, število stanj pa je večje celo pri MDKP, ki predstavlja slovar z 9 vnosi. Rezultati nakazujejo, da zastopanost pregibnih oblik močno vpliva na kompleksnost pridobljenega MDKP ter lahko vpliva na obrat trenda rasti velikosti MDKP.

Slika 4 prikazuje shematski prikaz dela MDKP, ki predstavlja končnice besed za slovar z 28 vnosi. Kompleksnost MDKP, ki predstavlja slovar z manj vnosi, je precej večja. Zato je smiselno, da so v slovarju, ki ga želimo realizirati s KP, prisotne vse možne pregibne oblike, saj si lahko v tem primeru leme, ki se enako pregibajo, del končnega pretvornika, ki pretvarja končnice, v celoti delijo. Kompleksnost pri tem še vedno povečujejo besede oz. leme besed, ki imajo med pregibnimi oblikami kakšno izjemo, ki se pregiba nekoliko drugače. V našem izmišljenem slovarju je to lema *zasp*, katere dve pregibni obliki imata nekoliko drugačno končnico, in sicer končnico *al* ter *ali* namesto *il* ter *ili*.



Slika 4: Del MDKP, ki predstavlja celoten slovar z vsemi 28 vnosi. Prikazan je le del, ki pretvarja končnice vnosov.

MDKP sprejme samo vnose, ki so vsebovani v slovarju. Če za določeno aplikacijo tako stroga zahteva ni potrebna in je dovolj, da MDKP sprejme vse vnose iz slovarja, ga lahko naprej poenostavimo. Še enostavnejšo obliko bi namreč dobili, če bi za vse štiri leme iz slovarja obstajale pregibne oblike za vseh 9 možnih končnic. V slovar lahko tako dodamo dodatne vnose in sicer vnose z lemami *potop*, *osmod*, *natoč* ter končnicama *al* ter *ali*, ter vnosa z lemo *zasp* in končnicama *il* ter *ili*. Pridobljeni slovar ima tako 36 vnosov, MDKP pa se poenostavi na 23 stanj in 30 prehodov.

4 Končni super pretvornik (KSP)

V prejšnjem razdelku smo pokazali, da lahko s pomočjo dodatnih, izbranih slovarskih vnosov v slovar zmanjšamo kompleksnost MDKP. Problem predstavlja iskanje takšnih slovarskih vnosov, ki bi zmanjšali kompleksnost, še posebej v primeru realnih slovarjev, kot so npr. slovarji izgovarjav, ki so prvič večji, drugič pa se ključ in vrednost posameznih slovarskih vnosov razlikujeta, s čimer je iskanje primernih slovarskih vnosov težja naloga.
Problema smo se zato lotili na drugačen način, in sicer tako, da smo združevali določena stanja, pri čemer smo želeli zadostiti naslednjima dvema pogojema:

- Pridobljeni KP mora ostati determinističen.
- Pridobljeni KP mora sprejemati vse ključe prvotnega slovarja ter za sprejete ključe oddati pravilne pripadajoče vrednosti.

Tako smo lahko združevali samo stanja, ki so imela določene lastnosti. Takšna stanja smo poimenovali združljiva stanja. Dve stanji sta združljivi, če zadoščata naslednjim pogojem:

- Če je eno od stanj končno stanje, stanji ne smeta imeti izhodnih prehodov s praznimi vhodnimi simboli oz. ε simboli. Rezultat združevanja takšnih stanj je lahko nedeterministični KP.
- Stanji nimata izhodnih prehodov z enakimi vhodnimi simboli ter različnimi izhodnimi simboli.
- Stanji nimata izhodnih prehodov z enakimi vhodnimi simboli ter enakimi izhodnimi simboli, ki prehajajo v različna naslednja stanja, ki so nezdružljiva.

Da bi lahko določili združljiva stanja, je potrebno preveriti zgornje pogoje, kar pa je v praksi lahko problematično, saj je preverjanje združljivosti stanj zaradi rekurzivnosti, ki je lahko ciklična, zahtevno. V ta namen smo zadnji pogoj poenostavili:

- Stanji nimata izhodnih prehodov z enakimi vhodnimi simboli ter enakimi izhodnimi simboli, ki prehajajo v različna naslednja stanja.

Zaradi poenostavitve pogoja za združljivost stanj nekaterih združljivih stanj nismo mogli zaznati. KSP smo zgradili tako, da smo najprej zgradili MDKP, nato pa smo nadalje združili vsa stanja, ki so združljiva. Za vsako stanje je bilo potrebno preveriti, ali je združljivo s katerim koli drugim stanjem. Ker nekatera stanja postanejo združljiva šele, ko združimo neka druga stanja, je bilo potrebno to storiti v več iteracijah.

5 Predstavitev slovarja izgovarjav s končnimi super pretvorniki KSP

Za razširjeni slovar izgovarjav iz razdelka 3 smo najprej zgradili MDKP s pomočjo odprtokodnega orodja OpenFST (Cyril at al., 2007), nato pa smo s postopkom, ki smo ga opisali v razdelku 4, zgradili še KSP. Tabela 2 prikazuje število stanj in prehodov MDKP in KSP.

Tabela 2: Zmanjšanje števila stanj in prehodov pri gradnji KSP iz MDKP.

		MDKP	KSP	Sprememba
1 izhodni simbol	*Stanja*	246.262	186.476	24.3%
	Prehodi	556.723	441.234	20.7%

Opazimo lahko, da smo velikost KSP v primerjavi z velikostjo MDKP uspeli zmanjšati za več kot 20%.

Čeprav lahko s KSP vnose v slovarju predstavimo z manjšim KP kot v primeru MDKP, pri tem izgubimo informacijo o tem, katere besede so vsebovane v slovarju. Tako se lahko zgodi, da KSP sprejme določeno besedo, ki je slovnično pravilna, vendar ni bila vsebovana v slovarju. V tem primeru je lahko oddan alofonski prepis napačen.

6 Zaključek

V članku je predstavljen nov tip KP, ki smo jih poimenovali končni super pretvorniki (KSP), ki poleg želenih besed sprejemajo še nekatere druge z namenom, da lahko pretvorbo želenih besed predstavimo bolj kompaktno.

Pokazali smo, da lahko pri predstavitvi slovarja izgovarjav s pomočjo KSP število stanj in prehodov zmanjšamo za več kot 20%, ko so za vsebovane leme v slovarju izgovarjav prisotne tudi vse pripadajoče pregibne oblike besed.

Ker KSP sprejemajo še druge, neznane besede, za katere lahko oddajo napačen izhodni niz, so KSP uporabni predvsem v aplikacijah, kje ne potrebujemo informacije o tem, katere besede so vsebovane v KP ampak le informacijo o pravilni pretvorbi danih besed oz. besed, iz katerih smo zgradili KSP.

Zahvala

Razvojno raziskovalno delo je delno sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije v sklopu aplikativnega raziskovalnega projekta OptiLEX (L7-9406).

LITERATURA IN VIRI

[1] Cyril A., Michael R., Johan S., Wojciech S., Mohri M., 2007. OpenFst: A General and Efficient Weighted Finite-State Transducer Library. Proceedings of the 12th International Conference on Implementation and Application of Automata (CIAA 2007). Lecture Notes in Computer Science, Prague, Springer-Verlag, Heidelberg, Germany, 4783: 11-23.

[2] Daciuk J., Piskorski J., Ristov S., 2011. Natural Language Dictionaries Implemented as Finite Automata. Scientific Applications of Language Methods. London: Imperial College Press, World Scientific Publishing.

[3] Golob Ž., 2009. Samodejno določanje mesta besednega naglasa pri sintezi slovenskega govora. Diplomsko delo, Fakulteta za elektrotehniko v Ljubljani.

[4] Golob Ž., Žganec-Gros J., Žganec M., Vesnicer B., Dobrišek S., 2012. FST-Based Pronunciation Lexicon Compression for Speech Engines. International Journal of advanced robotic systems, 9: 2012.

[5] Golob Ž., Žganec-Gros J., Štruc, V., Mihelič, F., Dobrišek S, 2016. A composition algorithm of compact finite-state super transducers for grapheme-to-phoneme conversion. Proceedings of the *Text, speech and dialogue Conference*, Brno, Czech Republic, September 12-16, 2016. Switzerland: Springer, str. 375-382, Lecture notes in artificial intelligence, 2016.

[6] Gros J., Mihelič F., 1999. Acquisition of an Extensive Rule Set for Slovene Grapheme-to-Allophone Transcription. Proceedings 6th European Conference on Speech Communication and Technology. September 5–9. 1999. Eurospeech 1999. Budapest, 5: 2075–2078.

[7] Hellis T., 2004. On minimality and size reduction of one-tape and multitape finite automata. Doktorska disertacija.

[8] Lucchesi C., Kowaltowski T., 1993. Applications of Finite Automata Representing Large Vocabularies. *Software-Practice & Experience,* 23: 15-30.

[9] Mohri M., 1994. Compact Representations by Finite-State Transducers. 32nd Meeting of the Association for Computational Linguistics (ACL '94). Proceedings of the Conference. Las Cruces. NM, pp. 204–209.

[10] Mohri M., 1997. Finite-State Transducers in Language and Speech Processing. *Computational Linguistics,* 33: 269–311.

[11] Ristov S., 2005. LZ Trie and Dictionary Compression. Jurnual Software-Practice & Experience, pp. 445–465.

[12] Žganec-Gros J., Cvetko-Orešnik V., Jakopin P., 2006. SI-Pron Pronunciation Lexicon: A New Language Resource for Slovenian. Informatica, 30: 447–452.

The Fundamentals of Sound Field Reproduction Using a Higher Order Ambisonics System

Rok Prislan*
rok.prislan@innorenew.eu
InnoRenew CoE
Livade 6, SI-6310, Izola, Slovenia

ABSTRACT

Conventional sound recording methods are based on recording the sound pressure level with a microphone which is after some signal processing reproduced by loudspeakers. In spatial audio, more than one microphone and loudspeaker are required to provide the sound source location information to the listener. Several spatial audio formats have been developed and some have successfully entered our homes, such as the the multichannel 5.1 surround system. Among spatial audio formats, Ambisonics stands out due to its capability of capturing and reproducing the whole sound field and is not limited to predefined loudspeaker setups. In the paper, the InnoRenew CoE's Ambisonics system is introduced and some of its underlying principles are explained. Furthermore, practical examples of the use of Ambisonics, also in relation to Virtual reality applications, are presented.

KEYWORDS

higher order Ambisonics, sound field reproduction

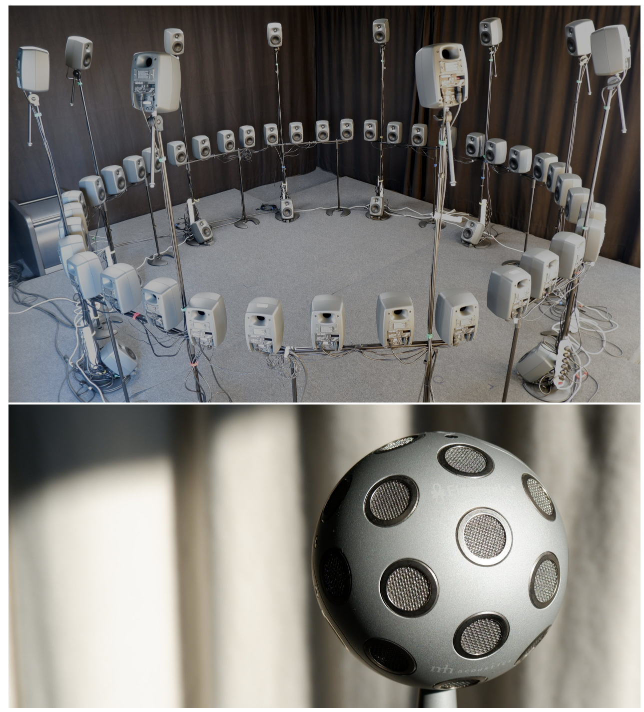


Figure 1: The higher order Ambisonics reproduction system with 64 loudspeakers (top) and the Ambisonics microphone [2] (bottom) which are part of the InnoRenew CoE's acoustic laboratory equipment.

1 INTRODUCTION

Michael Gerzon [1] invented Ambisonics in the 1970s, and since it has mainly been a research topic in acoustics. It's higher order version was developed twenty years later but only recently it has become a commercially available recording system [2]. Currently, more and more user applications of Ambisonics are emerging since Ambisonicsis is being positioned as the audio framework of choice for virtual reality [3, 4].

The acoustic laboratory of InnoRenew CoE has currently been equipped with a higher order Ambisonics system. The system is composed of a 32 channel microphone [2], a set of 64 full range loudspeakers, a dedicated low frequency loudspeaker, all the required AD/DA converters and accessories, such as stands and cables. The equipment in shown on Figure 1.

The system will be used for perceptual acoustic experiments, mainly by exposing test subjects to different acoustic conditions and investigating their response. In fact, room acoustic conditions are essential for a healthy and creative working environment – one of the important research topics at InnoRenew CoE. Another use of Ambisonics is in combination with virtual reality systems (e.g. [7]) that can provide a multi-sensoric immersion experience to users.

2 RECORDING AND ENCODING

Ambisonics is a method of recording and reproducing a sound field and preserving its directional properties. The signal is coded, which is different in comparison with traditional multichannel audio formats (e.g., stereo, and 5.1 surround). In those, each channel contains the signal corresponding to a loudspeaker while in Ambisonics each channel contains derivatives of the pressure field. The encoded signals are known as B format.

In Ambisonics we record with several microphones spherically arranged on a (virtual) sphere. Summing properly weighted signals from each microphone is equivalent to recording with a microphone of a certain directional characteristic. Such processing is the basis of Ambisonics encoding [2], in which case the chosen directional patterns correspond to spherical harmonic functions (see figure 2).

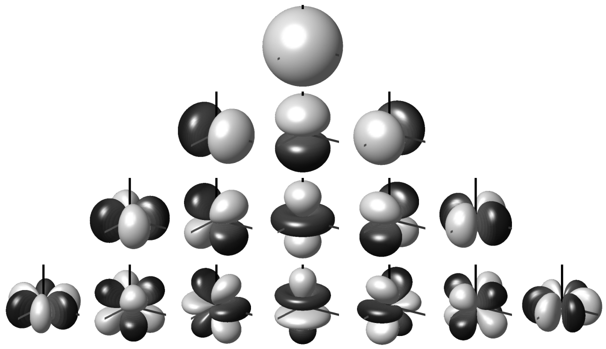


Figure 2: Polar patterns of spherical harmonics $Y_n^m(\theta,\varphi)$ of zero, first, second, third and fourth order (from top to bottom) (figure from [5]).

Spherical harmonic functions are grouped by their order number n and particular coefficient $m = -n, ...n$. Mathematically, each spherical harmonic corresponds to the angular portion of the solution of the wave equation. This way it is possible to capture the whole sound filed as it can be, in fact, decomposed into spherical harmonic functions

$$p(k, \mathbf{r}, \theta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi i^n j_n(kr) A_{n,m} Y_n^m(\theta, \varphi) \qquad (1)$$

where φ and θ are the azimuth and elevation, $\mathbf{r}$ is the spatial coordinate and k is the wavenumber.

The general idea of a higher order Ambisonics encoding is to record sound with directionality patterns that correspond to polar patterns of spherical harmonics. As such, it is possible to encode the sound field in form of spherical harmonic decomposition factors instead of the sound pressure level at each microphone position.

The maximum order N at which we perform the expansion defines the order of the Ambisonic system. Each order contains $2N + 1$ channels, meaning that in total the ambisonics system of order N has $(N + 1)^2$ channels that have to be stored. Increasing the order to which the decomposition is done improves the directionality of the recording.

An important limiting factor for increasing the Ambisonic order is the number of microphones positioned on the sphere: the pressure is discretely sampled, which leads to artifacts, such as aliasing. Issues related to low frequency noise and several other technical limitations have been studied [3]. Generally, increasing the number of microphones is favored, although this obviously increases the cost of the system.

It is important to understand that the B format encoded signals can be as well manipulated with proper signal processing. For example, the sound field can be easily rotated for a certain angle, and it is also possible to focus to a certain direction of the sound field [6].

3 REPRODUCING THE SOUND FILED

The biggest advantage of Ambisonics over conventional multichannel spatial audio techniques (e.g. stereo, 5.1 and 7.1 surround) that consider fixed loudspeakers position is independence on the loudspeaker setup. In Ambisonics, the decoding from the B format takes into account the actual position of the available loudspeakers, which can be arbitrary chosen. Nevertheless, a high number of loudspeakers spatially distributed around the listener are required to provide a full and precise spatial impression.

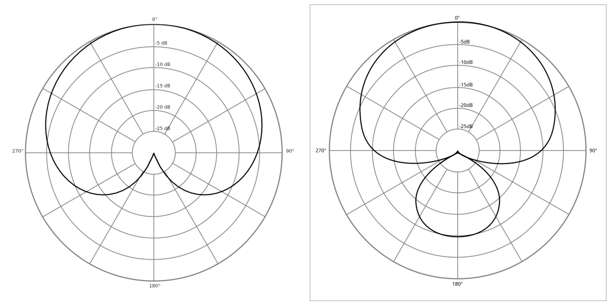


Figure 3: Example of a cardioid (left) and supercardioid (right) microphone polar pattern (figure from [8]).

The number of loudspeakers required is as well dependent on the order of the system. The $N-$th order requires a minimum $(N+1)^2$ loudspeakers, meaning that 9 loudspeakers are required for the 2nd order, 16 for the 3rd and 25 for the 4th.

There are several strategies for decoding the B format to be reproduced on a setup of loudspeakers. The basic idea is to directionally filter the recorded signals by virtual microphones pointing in the direction of each loudspeaker.

Setting the proper directionality patterns (see Fig 3) is the important part of the decoding process. In a regular layout, the signal emitted by a loudspeaker is the same as it would be recorded by a supercardioid microphone pointing towards that direction [6]. This means almost all loudspeakers emit sound at the same time, and for a given sound source position, loudspeakers in the opposite direction emit in opposite phase.

4 THE AMBISONICS SYSTEM IN USE

Ambisonics systems are an useful research tool in acoustics, mainly because they enable to reproduce sound emitted by sources together with the acoustic environment in which they are located. An important example of such use are the investigations carried out by Tapio Lokki [9] with his group who have been investigating perceptually relevant acoustic properties of concert halls. In their research, listeners have been asked about their preferences about the acoustics of different concert halls in which the same orchestra was performing. As an individual's acoustic memory is strongly affected by the time that has passed since each concert experience, it is required for such research to migrate the listener and orchestra between concert halls immediately. This can be achieved by an Ambisonics system in which recordings can be switched by a push of a button.

Figure 4: Photo of a listener in the Ambisonics loudspeakers ring at the InnoRenew CoE's Acoustic lab. The control over the system and perceptual response is based on a tablet PC as an interface.

Currently at InnoRenew CoE, we are setting up the Ambisonics system for the listener to rate different acoustic environments. The research is not limited to a specific environment type, such as concert halls, but includes acoustic environments to which we are exposed on a daily basis (commonly referred to as soundscape [12]). The recording will be performed on several different locations that include noisy and pleasant environments, such as high-traffic roads, busy workspaces and nature.

The interaction of the user with the system can be designed in various ways. Firstly, we are relying on a tablet PC as shown in Fig. 4. Using the tablet, the playback is controlled and the response from individuals is gathered. The system can be upgraded with more advanced response tracking options, such as performing eye-tracking or tracking the electrodermal activity of the test subject.

Spatial sound can be incorporated into virtual reality (VR) interfaces, such as VR headsets. The most accessible approach is to use headphones for which the signals have to be processed based on Head-related transfer functions [10]. The main drawback in this case is that wearing headphones is not natural to users and can produce discomfort. It is well known [11] that the listener does not localize the sound source as being external, but rather positions it in between the ears. This phenomenon of using headphones is known as lateralization of sound sources [11].

Generally, the relative position/orientation of the sound source in relation to the listener's ears changes over time, meaning that Head-related transfer functions applied to process the audio content have to adopt accordingly. Therefore, when using headphones in VR head tracking and real time audio processing are required.

In this perspective, the use of Ambisonics advantageous as the full sound field is reproduced and the listener can freely rotate his head while localization clues are correctly perceived. Additionally, in Ambisonics the ears are free from wearable equipment, which is a more natural condition for the user.

A relevant use of Ambisonics in relation to VR is also recording the sound field using an Ambisonics microphone and reproducing it over headphones instead of an Ambisonics reproduction system composed of a high number of loudspeakers. In fact, the B format encoded signals can be processed for a binaural playback for any arbitrarily chosen head rotation. Recently, many commercial second order Ambisonics microphones containing four microphones have become available on the market together with dedicated digital audio work station plug-ins for binaural decoding.

5 ACKNOWLEDGMENTS

The author gratefully acknowledges the European Commission for funding the InnoRenew project (Grant Agreement #739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding from the Republic of Slovenia and the European Union's European Regional Development Funds).

REFERENCES

[1] Michael A Gerzon. *Periphony: With-height sound reproduction.* In: Journal of the Audio Engineering Society 21.1 (1973), pp. 2–10.
[2] Eigenmike — mh acoustics LLC, url: https://mhacoustics.com/products (visited on 16/09/2020).
[3] F. Zotter and M. Frank: *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, SpringerOpen, 2019
[4] S. Sherbourne et al. Ambisonics and VR/360 Audio in Pro Tools url: http://www.avidblogs.com/ambisonics-vr360-audio-pro-tools-hd/ (visited on 16/09/2020)
[5] Ambisonics — Wikipedia, The Free Encyclopedia. url: http://en.wikipedia.org/w/index.php?title=Ambisonics&oldid=656474391 (visited on 16/09/2020).
[6] D. Arteaga, *Introduction to Ambisonics.* url: https://www.researchgate.net/publication/280010078_Introduction_to_Ambisonics (visited on 16/09/2020)
[7] Virtual Acoustics — Institute of Technical Acoustics RWTH Aachen University url: http://virtualacoustics.org/ (visited on 16/09/2020)
[8] Microphone — Wikipedia, The Free Encyclopedia. url: https://commons.wikimedia.org/w/index.php?curid=50230608 (visited on 16/09/2020).
[9] T. Lokki, *Tasting music like wine: Sensory evaluation of concert halls,* Physics Today, 67(1), 27 - 32, 2014, doi: 10.1063/PT.3.2242
[10] Head-related transfer function — Wikipedia, The Free Encyclopedia. url: https://en.wikipedia.org/wiki/Head-related_transfer_function (visited on 24/09/2020).
[11] Yost W.A., Hafter E.R. *Lateralization.* In: Yost W.A., Gourevitch G. (eds) *Directional Hearing.* Proceedings in Life Sciences. Springer, New York, NY (1987).
[12] Soundscape — Wikipedia, The Free Encyclopedia. url: https://en.wikipedia.org/wiki/Soundscape (visited on 24/09/2020).

The use of eCare services among informal carers of older people and psychological outcomes of their use

Kaja Smole-Orehek
kaja.smole-orehek@fdv.uni-lj.si
University of Ljubljana, Faculty of Social Sciences
Ljubljana, Slovenija

Vesna Dolničar
vesna.dolnicar@fdv.uni-lj.si
University of Ljubljana, Faculty of Social Sciences
Ljubljana, Slovenija

Simona Hvalič-Touzery
simona.hvalic-touzery@fdv.uni-lj.si
University of Ljubljana, Faculty of Social Sciences
Ljubljana, Slovenija

ABSTRACT

With increasing age and longevity, the need for informal care will increase significantly in the coming decades. The use of eCare services has potential benefits in meeting some of informal carer's needs. However, there is only a limited understanding of the psychological outcomes of using eCare services for informal carers of older people. The aim of this study is to identify positive and negative psychological outcomes of the use of eCare services for employed informal carers of older people, and to review the psychological outcomes of the use of different functionalities of eCare services. We have conducted a four-month intervention study among 22 dyads of informal carers and older people. The preliminary results showed a prevalent pattern of positive outcomes of eCare services for employed informal carers. Further research is needed on the relationship between the use of different functionalities, psychological outcomes and care situations.

KEYWORDS

psychological outcomes, employed informal carers, telecare, ageing in place

1 INTRODUCTION

The growing pressure on families to provide informal care, due to demographic aging of the population, leads to a search for new and innovative solutions to meet those challenges. An increasing attention is being paid to the role of technology and its potentials in supporting older people in their own homes and their informal carers. However, understanding the psychological outcomes of the use of eCare services is limited for informal carers of older people and even more so for working informal carers [1, 4, 6, 11, 14, 18].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia
© 2020 Copyright held by the owner/author(s).

Informal carers provide not just physical support, but also social and emotional, as well as making sure that the older people are safe and healthy, so informal care is very demanding and dynamic [17]. Informal carers are spending a lot of time at the home of the care receiver and the demands of providing care can be high, especially to those who are employed. Because of caring duties, some informal carers addressed different issues, such as work interruptions, absences and reduced productivity [3, 17, 20]. Combining employment and care is a challenge to many carers and it can have an influence of informal carers' physical health, social relationships, as well as the work situations [17]. Many are having troubles being understood by their employers or co-workers and to some the career opportunities cannot be obtained [17].

ECare services have a potential to address those needs, such as decrease the demand on carers and stress alleviation [1, 4, 11, 17]. Many studies are studying the link between psychological outcomes and the functionalities of eCare services, but these studies are very disease-specific, such as dementia [2, 12, 19]. Informal carers of people with dementia have specific needs, and these needs cannot be fully transferred to the needs of informal carers providing a different type of care, therefore there is a gap in understanding which functionalities of eCare services can help informal carers in general to better combine work and care. In addition, many studies examined different outcomes and models of eCare services use and acceptance among the older people [8, 15], living aside informal carers of older people.

The aim of the study was to fill this gap and to identify the positive and negative psychological outcomes of the use of eCare services on employed informal carers and to review the psychological outcomes of the use of different functionalities of eCare services.

2 METHODS

Study design

A four-month qualitative intervention study was performed in 2018 and 2019 in Central Slovenia region. In accordance with the aim of the study, surveys and interviews were conducted with employed informal carers only. The intervention study included a dyad of an informal carer and an older person (65+). A total of 26 dyads, including older care receivers and their primary informal carers, were recruited. The final sample included 22 dyads.

Apparatus

Older people had one of the two eCare equipment installed in their homes. Their informal carers used the mobile application, which allowed them to monitor certain activities in the older person's home and receive notifications in case of unexpected event. Both services tested had motion and door sensors, a pendant alarm, a smoke detector and a mobile application for the carers, with alarms in the form of push notifications and activity monitoring. The second service also offered two additional functionalities, a 24/7 call center and fall detection, and was used by 7 out of 22 carers.

Participants

Purposive sample was used. Eligibility criteria for informal carers were: (i) primary carers, (ii) family members of older person, (iii) providers of a long-term care to older person (providing at least 5 hours of help per week and minimum 1 year), (iv) owned a smartphone, (v) interested in study participation. Care receiver's eligibility criteria were as follows: (i) interested in study participation, (ii) old 65 years or more, (iii) need help with activities of daily living, (iv) use also formal care, (v) live alone in their own household.

Informal carers ranged in age from 35 to 67 (M = 53.9, SD = 7.56). More than half of them were female (n = 14). On average they provided 8.5 hours of care per week (SD = 12.15), on average for 6.1 years (SD = 5.89). A great majority of carers were care receiver's children (n = 20) and two were daughters-in-law.

Care receivers were on average 83 years old (SD = 6.04), ranging from 73 to 92 years. All but two were females and all but one fell in the last five years, with 14 of them needed medical assistance afterwards. Five of them were severely dependent, eight moderately dependent, eight slightly dependent and one needing some help only occasionally.

Procedure

During the intervention, quantitative (screening questionnaire) and qualitative (semi-structured interviews) data were collected, with qualitative methodology playing a fundamental role. The survey at the beginning of the intervention collected basic social, health, care and demographic data. Two semi-structured interviews per informal carer were then conducted (the first after 3 weeks of use and the fourth month), each lasting about one hour. They were asked about their caregiving situation, their experience with new technology, their use of the tested eCare services and the psychological outcomes of eCare services. The in-depth interviews were recorded and completely transcribed. Personal information was made anonymous. All participants received gift vouchers in recognition of their time and were not charged for their use of the eCare services. The study was approved by the Slovenian Commission for Medical Ethics (0120-193/2018/15).

Analysis

A descriptive analysis of the quantitative data, comprising 755 pages of transcribed interview recordings, was carried out. The qualitative data were subjected to a thematic analysis using the programme Atlas.ti 8 for qualitative data analysis. A structural coding was used. This is question-based code that "acts as a labelling and indexing device, allowing researchers to quickly access data likely to be relevant to a particular analysis from larger data set" [13, 16]. Deductive and inductive approaches were combined for data coding and analysis [7].

3 RESULTS

We examined psychological outcomes of five eCare services functionalities: motion detection on the App, Push notifications and alarms on the App, Emergency pendant, Smoke detector, Call center and Fall detector. The most frequently reported positive psychological outcome was reassurance, followed by peace of mind and reduced anxiety. In addition, the informal carers mentioned several other positive psychological outcomes of using eCare services, including an increased sense of control, less stress, the feeling of being less burdened, having positive feelings, a feeling of relief and satisfaction (Table 1).

	Push notifications and alarms on the app	Sensor-based motion detection on the app	Call centre assistance	Fall detector	Emergency pendant	Total
Reassurance	8	18	7	2	8	43
Peace of mind	8	12	4	1	6	31
Reduced anxiety	5	12	2	0	3	22
Reduced stress	3	6	1	0	1	11
Feeling less burdened	2	4	2	0	1	9
Positive feelings	0	2	1	0	1	4
A sense of relief	0	3	0	0	0	3
Satisfaction	0	0	1	0	1	2
Total	26	57	18	3	21	

Table 1: Positive psychological outcomes of eCare services use on the employed informal carers

"Yes, yes, that gives you the feeling of reassurance that they will inform you, if you are not around when he presses the button. And then you go on vacation or somewhere else, as I say, even if something would happen, you organize other family members to make an action." (Carer 15)
"I will say it, if I did not reach mum over the phone call because of her bad hearing, then I looked at this application and saw that mom is inside doing something. If something was wrong, there is also an option for an emergency pendant, which she could activate, right ... " (Carer 6)

In our study, informal carers recognized that the emergency pendant and sensor-based motion detection are two of the most helpful functionalities of eCare services, as they are the most common contributors to positive psychological outcomes. The most useful functionality for those who cared for people with severe disabilities was an emergency pendant, while sensor-based motion detection was more useful for those carers who cared for people with mild or moderate dependency.

The few negative psychological outcomes in our study were mostly caused by technical failure and false alarms, although some participants were less disturbed than others. The most frequently mentioned negative outcome was anxiety, followed by distrust and stress. Other reasons for negative psychological outcomes were mentioned: feelings of false security, invasion into older person's privacy, feelings of guilt because they do not help enough, increased worries because they know the everyday patterns of older person (Table 2).

"Hm, yes, it caused me more stress personally because I was worried, under other circumstances I would not do that. Under other circumstances, I would not think about whether she was still cooking or not." (Carer 2)

	Push notifications and alarms on the app	Sensor-based motion detection on the app	Call centre assistance	Fall detector	Emergency pendant	Total
Anxiety	6	3	1	0	0	10
Hesitant	0	0	7	0	0	7
Distrust	1	3	1	1	0	6
Stress	2	2	0	0	0	4
Feeling burdened	1	3	0	0	0	4
Lack of relief	0	1	0	0	2	3
Doubts	0	1	0	0	2	3
Discomfort	0	2	0	0	0	2
Less peace of mind	0	2	0	0	0	2
Additional problem	1	1	0	0	0	2
No reduced burden	0	1	0	0	0	1
Feeling a moral obligation	0	1	0	0	0	1
Sense of guilt	0	1	0	0	0	1
Bothered	1	0	0	0	0	1
Unpleasant feeling	1	0	0	0	0	1
Total	13	21	9	1	4	

Table 2: Negative psychological outcomes of eCare services use on the employed informal carers

"Even more, because she does not want to wear this neckless, then it seems to me to be useless. You do not need it for anything, it will not be very functional, because then it will not matter if she only has a mobile phone." (Carer 2)

In the present study, all call center users mentioned positive psychological outcomes in relation to it. They mostly felt reassured by their service. However, a few participants who did not have access to the call center service felt reluctant to use it because they said that they might not have enough information about the older person to be able to react well, that they would not feel comfortable talking to a "stranger" and that their situation was too specific for a call center to be helpful. They were worried that the call center would approach to older person too technically rather than attentively.

"... I know, even when mom falls, she is always very confused. It is much easier when she says, "daughter, I fell," opposing to explain it to them. Well, "lady, this and that". I do not know, I am absolutely for an option that one of the family members has it." (Carer 4)

"It would not help much. It would not help us, because if mum is alone when she falls (...). I can assume that she is able to say something on her own, maybe not, right." (Carer 8)

For some carers, eCare services contributed to their ability to be in paid employment and was useful in reconciling care and work obligations. They mentioned that eCare services supported their work and increased their labor productivity by making it easier for them to concentrate on their work.

"Yes, I can concentrate at work. I do not have to think about what if ... It helps, and it relieves you, but still, if she falls, and her phone is ten feet away and she cannot use it, then it's useless ..." (Carer 9)

4 DISCUSSION

Our study yielded several important findings. We found that positive psychological outcomes of eCare services use for employed informal carers were much more common than negative ones. This finding supports the findings of previous limited studies of informal carers [1, 9, 14]. Despite the prevalence of positive psychological outcomes, the negative should not be ignored. In particular studies show that unreliable and/or inappropriate technology, which in our study was the main cause of negative psychological outcomes, can be harmful to both older people and their informal carers [5, 10]. In addition, a difference was also observed in the perceived usefulness of individual functionalities in relation to the degree of dependence. However, due to the small number of participants in different dependency groups, further empirical and conceptual studies are needed. Our study also confirmed the complex relationship between the functionalities of eCare services and the psychological outcomes for employed informal carers.

We also demonstrated that reassurance was the most frequently identified positive psychological outcome. It was mainly related to sensor-based motion detection on the application, the possibility of monitoring the activities of an older person from distance, e.g. to verify that he or she is moving around home safely. In addition, employed informal carers reported that reassurance makes it easier for them to go on business trips, work and concentrate on their work, as they are notified in case of an emergency. From this, we can conclude that eCare services can provide opportunities for employed informal carers of older people to reconcile work and care responsibilities.

This study examined under-researched aspect of eCare use in relation to informal care of older people. The methodology used allowed a detailed account of the experiences of employed informal carers' with eCare as well as their perceptions of it. However, there are some limitations to this study. The first is the duration of the intervention. When conducting an intervention study to investigate the detection and vigilance of a potentially harmful event, a longer duration of the intervention is usually advisable, but we were limited in time and resources. In addition, the incidence of a harmful or unexpected event during the testing phase in our study was low, so many participants had no real experience with the support and protocols for using eCare services. Moreover, one of the eCare services tested was still in the testing phase during the intervention study, so that several false alarms occurred, especially at the beginning of the study.

5 CONCLUSIONS

Our study confirmed the potential of eCare services to address challenges related to long-term care provision. There are many challenges that Slovenian society needs to address in order to realize the full potential of eCare technologies: (i) Public authorities must recognize the role and caring demands of informal carers and provide them with much needed support as soon as possible. (ii) Policy makers should promote a policy framework for the creation of eCare services for carers and beyond [19]. (iii) Affordable and accessible eCare services must be made available to informal carers and older people [18, 19]. At the same time, we must increase their acceptance of such technologies. Therefore, the design and usability of these technologies should be adapted and personalized to the needs of informal carers [2, 18]. End users should therefore be involved in the test phases [5, 18].

ACKNOWLEDGMENTS

The authors acknowledge the projects (Smart ICT Solutions for Active and Healthy Ageing: Integrating Informal eCare Services in Slovenia, ID L5-7626, Factors impacting intention to use smart technology enabled care services among family carers of older people in the context of long-distance care, ID J5-1785, Programme Internet research, P5-0399) were financially supported by the Slovenian Research Agency.

REFERENCES

[1] Stefan Andersson, Christen Erlingsson, Lennart Magnusson, and Elizabeth Hanson. 2017. Information and communication technology-mediated support for working carers of older family members: an integrative literature review. *IJCC* 1, 2 (2017), 247–273.

[2] Laura Block, Andrea Gilmore-Bykovskyi, Anna Jolliff, Shannon Mullen, and Nicole E Werner. 2020. Exploring dementia family caregivers' everyday use and appraisal of technological supports. *Geriatric Nursing* (2020).

[3] Jill I Cameron, Rene-Louise Franche, Angela M Cheung, and Donna E Stewart. 2002. Lifestyle interference and emotional distress in family caregivers of advanced cancer patients. *Cancer* 94, 2 (2002), 521–527.

[4] Stephanie Carretero, James Stewart, and Clara Centeno. 2015. Information and communication technologies for informal carers and paid assistants: benefits from micro-, meso-, and macro-levels. *European Journal of Ageing* 12, 2 (2015), 163–173.

[5] Sara J Czaja. 2016. Long-term care services and support systems for older adults: The role of technology. *Am Psychol* 71, 4 (2016), 294.

[6] Anna Davies, Lorna Rixon, and Stanton Newman. 2013. Systematic review of the effects of telecare provided for a person with social care needs on outcomes for their informal carers. *Health & social care in the community* 21, 6 (2013), 582–597.

[7] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *Int. J. Qual. Methods* 5, 1 (2006), 80–92.

[8] Jean D Hallewell Haslwanter and Geraldine Fitzpatrick. 2018. The development of assistive systems to support older people: issues that affect success in practice. *Technologies* 6, 1 (2018), 2.

[9] Nat Harward. 2016. Caregivers & technology: What they want and need. http://www.aarp.org/content/dam/aarp/home-and-family/personal-technology/2016/04/Caregivers-and-Technology-AARP.pdf. Accessed: 2020-09-10.

[10] Helen Hawley-Hague, Elisabeth Boulton, Alex Hall, Klaus Pfeiffer, and Chris Todd. 2014. Older adults' perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. *International journal of medical informatics* 83, 6 (2014), 416–426.

[11] Kara Jarrold and Sue Yeandle. 2009. A weight off my mind: Exploring the impact and potential benefits of telecare for unpaid carers in Scotland. https://www.scie-socialcareonline.org.uk/a-weight-off-my-mind-exploring-the-impact-and-potential-benefits-of-telecare-for-unpaid-carers-in-scotland/r/a11G0000001825wIAA. Accessed: 2010-09-30.

[12] Guang Ying Mo, Renée K Biss, Laurie Poole, Bianca Stern, Karen Waite, and Kelly J Murphy. 2020. Technology Use among Family Caregivers of People with Dementia. *CAG/ACG* (2020), 1–13.

[13] Emily Namey, Greg Guest, Lucy Thairu, and Laura Johnson. 2008. *Data reduction techniques for large qualitative data sets.*

[14] Kaja Smole Orehek, Ines Kožuh, Andraž Petrovčič, Vesna Dolničar, Matjaž Debevc, and Simona Hvalič-Touzery. 2018. Psychological outcomes of eCare technologies on informal carers of older people. *IJIC* 18 (2018).

[15] Sebastiaan Theodorus Michaël Peek. 2017. *Understanding technology acceptance by older adults who are aging in place: A dynamic perspective.* Ph.D. Dissertation. Tilburg University.

[16] Johnny Saldaña. 2015. *The coding manual for qualitative researchers.* Sage.

[17] Alice Spann, Joana Vicente, Camille Allard, Mark Hawley, Marieke Spreeuwenberg, and Luc de Witte. 2020. Challenges of combining work and unpaid care, and solutions: A scoping review. *Health & Social Care in the Community* 28, 3 (2020), 699–715.

[18] Vimal Sriram, Crispin Jenkinson, and Michele Peters. 2019. Informal carers' experience of assistive technology use in dementia care at home: a systematic review. *BMC geriatrics* 19, 1 (2019), 160.

[19] Vimal Sriram, Crispin Jenkinson, and Michele Peters. 2020. Carers' experience of using assistive technology for dementia care at home: a qualitative study. *BMJ open* 10, 3 (2020), e034460.

[20] Birgit Trukeschitz, Ulrike Schneider, Richard Mühlmann, and Ivo Ponocny. 2013. Informal eldercare and work-related strain. *J GERONTOL B-PSYCHOL* 68, 2 (2013), 257–267.

Indeks avtorjev / Author index

IS 2020

Interakcija človek-računalnik v informacijski družbi

Human-Computer Interaction in Information Society

Veljko Pejović, Matjaž Kljun, Vida Groznik, Domen Šoberl, Klen Čopič Pucihar, Bojan Blažica, Jure Žabkar, Matevž Pesek, Jože Guna, Simon Kolmanič