PERSONALIZING LINEAR RADIO:
MODEL AND USER EVALUATION

Paolo Casagranda[1], Francesco Russo[1], Raffaele Teraoni Prioletti[1]

[1]Rai, Research & Technology Innovation Centre, Italy

ABSTRACT

The paper introduces radio personalization, as part of the strategy to keep radio relevant, taking into account new available technologies and audiences changing their media consumption habits. A novel model for proactive context-aware radio personalization based on the hybrid content radio framework is discussed, and a prototype based on the model is described. Finally, a user evaluation of the personalized radio service is described, and the results are discussed, showing a path of evolution for future linear radio.

INTRODUCTION

Radio has been a strong and popular medium for almost a century. In the last years, technology has created the foundations to quickly build and distribute a broader range of audio services, mainly recommendation-based music playlists, and new players are challenging traditional broadcast radio, acting as competitors or content distributors and possibly as gatekeepers. On the other hand, traditional radio is slowly losing ground, especially among young audiences. Tailoring radio for each listener and mixing editorial scheduling with recommended content in a seamless way is both one of the possible strategies to keep radio relevant, and an evolution of the radio service itself to better fulfil the goals of public service.

After an overview of a few relevant projects in the area of radio personalization, we propose a model for radio personalization services, based on the hybrid content radio (HCR) framework and leveraging on context-aware and content-based recommendations. Recommendations are based on user behaviour and content semantics on the one side, and available listening time, content complexity and channel cost, on the other.

A prototype inspired by the model has been developed to validate the concept, following the framework presented in [1]. Specifically, the purpose was to measure user propensity to switch between linear and on-demand content and user reaction towards automatic proactive recommendations. A first panel of more than 200 volunteers briefly tested the prototype and participated in a survey. Then, a second panel of 50 volunteers was selected to evaluate interactive and proactive personalization using a customized prototype installed on their smartphones for two weeks. Very positive feedbacks on content-search and skip functions emerges, while the feedbacks towards proactive personalization suggest that further investigation may be required.

RADIO PERSONALIZATION: ONGOING ACTIVITIES

In 2005 Google Labs a project on personalized search went in production, and in 2009 it was made available to signed-out users too: the search for personal relevance reached the masses. From 2008 Spotify began offering personalized music playlists, and at the end of 2015 Capital XTRA announced MyCapital XTRA App, with a content skip function allowing access to a list of alternative content, making live radio more flexible.

From 2014, Public Service Media radio stations also began innovating their linear radio services. Listeners were allowed to choose also personally recommended content, see the Arena app from YLE in Finland, or BBC radio app in the UK. BBC proposed in 2014 the Object-Based Broadcasting framework [2], with the purpose of adapting the audio stream to the listening device and context. Rai together with other European broadcasters proposed the HCR framework to mix linear radio and audio on-demand (AOD, see e.g. [3]) leveraging content recommendations. And the Swedish SR innovated their service working on audio *atomization*: AOD for the news was produced with fine-grained metadata, allowing access to each news element independently. Also NPO recently started an innovation project, NPO Radio1 To Go, especially targeting commuters using cars and train: their available time is used to provide the news bulletin as a first item, and then a personalized playlist.

Research is also shared within the European Broadcasting Union Strategic Programme on Radio, which kicked-off in 2017 a Project Group on Radio Personalization: the purpose is to share know-how and to find the key aspects of personalization for Public Service Media.

HYBRID CONTENT RADIO

Hybrid content radio (HCR) aims at making linear broadcast radio more flexible, fostering the mix of traditional linear radio content with AOD, leveraging on content replacement, time-shifting and context-aware content recommendations [1]. The replacement can be triggered in a number of ways: a feedback from the user, a skip command or a recommender-driven proactive replacement.

Figure 1 - UI of the prototype app

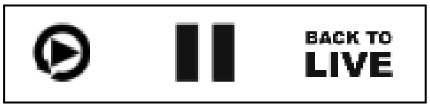
Metadata following the ETSI standards of the RadioDNS Project, see e.g. [4], technically allows content replacement in HCR. The recommender system used in the HCR service provides contextually-relevant alternative audio content [5]. An example use case involves in-car listeners: after having learned the listener's preferences, the trip duration, the road conditions and geometry and other information, we can increase the relevance of the content, e.g. selecting appropriate content duration or proposing the news at the beginning of the trip. Using linear radio as the basic building block for personalized radio has some advantages:

- the relevance of the content increases, enriching the listeners' experience, while listeners can keep a link to their favourite radio station
- the efficiency of content delivery can be optimized, if the device allows using a broadcast technology to receive the audio from the broadcast channel

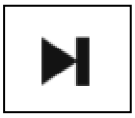
USER EXPERIENCE ELEMENTS

We developed a prototype app for Android smartphones to validate the concept and include previous user feedbacks: Figure 1 shows the prototype app main view. The prototype included interface elements allowing to control the mix between linear radio and AOD *manually.* Moreover, the prototype allowed proactively recommended content to be *automatically* played: the recommender suggested both the content and time to deliver it.

In the following paragraph we describe the more important interface elements allowing to *manually* trigger the stream personalization.



The listener can play, pause and restart the programme (start from the beginning – see the figure on the left). The listener can always go back to the live programme.



The interface allows the listener to "skip" current programme and access a personalized list of recommended AODs (see the figure on the left).

A search button (see the figure) performed a full-text search on the audio content, based on the metadata and the automatically extracted transcript.

If the listener selects an element from the search results list, the player starts the AOD. The search keyword occurrences ("mondiali" in Figure 2) then appear on the AOD timeline and the listener can navigate them.

The "similar clips" list proposes for each station or AOD being played a list of similar, recommended AODs, see Figure 3. The recommendation algorithm leverages content-based semantic similarity, as described below.

The *automatic* personalization mechanism is based on a context-aware proactive recommender, as described in [6]. A proactive recommender suggests both the item to be delivered *and the time to deliver it*. Proactive recommenders have been recently gaining popularity due to the applicability, especially for mobile applications, smart speaker based

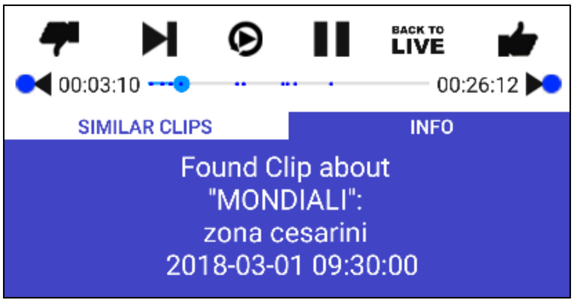


Figure 2 – AOD timeline navigation

Figure 3 – Semantically similar AODs

services and the in-car environment. Where interaction should be limited, proactive recommendations can be very useful. The automotive environment is a relevant example: driver's distraction should be reduced for safety reasons [7], and proactive recommendations can take advantage of the context [8].

PROACTIVE CONTEXT-AWARE PERSONALIZATION: THE SCHEDULE BUDGET MODEL

Interactive recommendations usually follow widely studied models and algorithms, for example content-based filtering, see [9]. For proactive recommendations an original model, the Schedule Budget Model, was chosen, taking into account some of the most representative constraints for content recommendations, see the details in [6]. In the following discussion a *media object* is an AOD, but it could be also an audio-visual or a web page. Each media object is modelled as an entity with a value represented by the relevance for a listener, and some cost properties that are constrained by budget variables. Relevance depends on the listener, the media object and of course on the listening context. We limited the analysis to a few of the cost properties: available time, attention and channel cost. We will show the meaning of each property referring to the relevant use-case of in-car listening.

Media Object Relevance. The relevance for each listener of media objects was estimated. For the prototype we chose a content-based recommender described below.

Time budget. It is the time available to the listener. If we can estimate the available listening time, we can more precisely select the set of media objects to be recommended. If the listener is travelling in her car, an approximation of the available time is the trip duration: e.g. it is not effective suggesting a 60’ media object if the trip is only 20’ long.

Attention budget. It is the overall mental workload of the listener. When deciding if a media object has to be recommended, content complexity is constrained by the current mental workload of the listener. Users usually listen to radio while doing other: so radio listening becomes most of times a secondary activity while doing a primary activity, and it is different listening to a radio drama or to an audio book while sitting on a sofa or while driving across the city in the rush hours, as quantified by [7] on the cognitive distraction scale. The mental workload level of tasks performed in parallel has been evaluated, showing that human cognitive, perceptual and motor resources are limited, see e.g. [10] and the complexity of each task contributes to the overall mental workload level, while this is still a matter of active research [11]. In particular, when driving a car is the primary activity distraction and mental workload level become important [7]. The model includes the mental workload level of the listener, allowing to constrain it, taking into account both her current activity and the content complexity. Of course at the moment the calculations are based on estimates, and this simplification only applies to homogenous complexity trip segments: the first part of the trip is very relaxing and the second requires high attention levels, the model can be applied to the two segments separately. Listening radio while driving is a relevant use case: a practical implementation can for example avoid complex audio content during the rush hours in the city centre.

Channel budget. The distribution cost of the media object on a specified channel can be modelled as a constrained property. Content relevance can then be maximized taking into account the overall delivery cost. For example, if two channels are available, cheap (broadcast content) and expensive (AODs from the mobile network), a bound can be set to limit the usage of the expensive channel.

Other properties can be added. The model captures in a simple way the relevant properties of media objects and we will see that allow a straightforward solution using combinatorial optimization. The model can be solved as a *multi-dimensional 0-1 knapsack* *problem (MKP)*, see [12]. Formally, we have set of n media objects $\{o_1, o_2, \dots o_n\}$ that map on the available AODs, and for each user u each o_j has an associated relevance v_j and an i-*th* cost property value is a_{ij}. E.g. $i=0$ corresponds to the object duration and the time budget, and so on, see [13]. Being x_j a coefficient indicating if the object belongs or not to the solution, the problem is to find:

$$\max \sum_{j=1}^{n} v_j x_j \quad s.t. \sum_{j=1}^{n} a_{ij}\, x_j \leq W_i \qquad with\ i \in \{1, 2, \dots, m\}, \qquad x_j \in \{0,1\}$$

The multi-dimensional knapsack problem is NP-Hard and no efficient polynomial time approximation scheme (EPTAS) exists [12], however solutions can be easily found using other heuristics and techniques, like genetic algorithms. For our investigations, starting from a list of 60 candidate media objects relevant for the listener, we used a genetic algorithm with an initial population of 300, mutation probability of 0.05, crossing probability of 0.8, 20 generations, with elitist selection, see [20]. Genetic algorithms are an affordable solution for similar problems: in [21], He et al. show how evolutionary algorithms based heuristics can help solving even time-varying knapsack problems.

Finally, the set of x_j values gives the list of recommended media objects: relevant and satisfying the constraints.

Validation of the model. The model has been validated for the in-car listening use case. This is a relevant scenario as statistically in-car radio listening accounts for a large audience in most countries: in Italy about 70% of audience, in the average day. The idea of helping users with recommended content and information, and even with personalized interfaces, has been already proposed, see [8], [14]. For the in-car listening use case, the time budget was approximated with the duration of the trip, constraining the recommended content list duration. The validation assumed the trajectory could be predicted: this assumption was verified using density based clustering from the GPS locations to find staying points, and then simplifying the emerging geometry and statistically analysing the results. The driver's task complexity was approximated with road geometry: crossroads and curves increased the driver's required attention. Also, the channel cost was simulated requiring part of the audio content to be taken from the broadcast channel. Using the GeoLife trajectory dataset and a second dataset built by researchers for the purpose, and AODs from RadioRai and the One Million Song dataset, we simulated in-car listening with different trajectory types and durations, validating the model for proactive content recommendations, and showing that it could effectively create recommended content lists.

RECOMMENDATIONS LEVERAGING SEMANTIC ANALYSIS

Recommendations can usually take advantage of several algorithm classes: collaborative filtering, content-based filtering, knowledge based filtering, social based and others, see [9]. During the trial we only took advantage of content-based recommendations as we wanted to assess semantic similarity, to be used in the future especially for the cold-start problem arising when new AODs are available. *We represented AODs as text metadata and transcripts, so AOD classification naturally translates to text classification*, a deeply studied research area. In the prototype we evaluated a content-based algorithm leveraging Explicit Semantic Analysis (ESA), see [15], [16]. ESA allows to represent an unrestricted natural language text as a concept vector, that is a vector in a *high dimensional space of concepts* from a knowledge base. Specifically, the vector space base used for the representation has been taken from Wikipedia 2016. The dimension of the vector space is almost four hundreds of thousands (385231 concepts in the version in use) and each AOD can virtually use all of the concepts to be represented. However, each concept vector can be reduced to obtain a compromise between the precision of the representation and the AODs similarity computing time. From our evaluation on an Italian language set of 2500 AODs, we found that reducing the number of concepts representing each AOD to 10000 concepts at most takes about 13% of computation time and a 20% error on the similarity calculation; while a set of 50000 concepts at most could take 25% of time with a 10% error. The classification and recommendation process is organized using the following steps:

- We build a repository including all AODs that can be suggested to the users (about 2500 elements). The recommended AODs for a listener are extracted from this set. The repository is updated day by day, including new AODs from Radio Rai and removing old ones.
- The *transcript* of each new AOD is extracted using a trained Automatic Speech Recognizer trained on the Italian language and is represented as a vector of concepts from Wikipedia.
- Each user's feedback to a AOD is associated to the concept vector representing the AOD itself.
- Each user is represented as a linear combination of the concept vectors $ESA(o_j)$ of the AODs she evaluated, weighted according to the feedback type (explicit or implicit), in the following way: $\boldsymbol{ESA}(\boldsymbol{user}_i) = \sum_{\forall\ rated\ o_j} \boldsymbol{w}_{ij}\, \boldsymbol{ESA}(\boldsymbol{o}_j)$, where $user_i$ is the i-th user, w_{ij} is the weight related to the feedbacks given by the i-th user to the AOD o_j and $ESA(o_j)$ is the concept vector associated to o_j . In this way, each user is represented by a vector in the concepts vector space.
- For each user, a score is assigned to each AOD in the repository of candidates based on the similarity to $\boldsymbol{ESA}(\boldsymbol{user}_i)$, and the score assignment is used to create an ordered list of candidates. Recommended clips for a user are then taken from the ordered list, choosing the candidates with highest scores. To calculate similarity, we chose the *cosine metric*.

With thousands of audio clips to be processed, indexed and compared to find similarity, calculation time tends to be not negligible. Given a free text sentence, we found semantic similarity search in the AODs repository was time consuming (several seconds), so without further optimizations it didn't scale up for thousands of parallel queries from the audience. On the other hand*, ESA is very well suited to calculate similarities between a fixed set of AODs*, in the following way. Given a set of AODs S_n with n elements, we calculate the matrix of mutual similarities M_s, where each element is $m_{ij} = \boldsymbol{sim}(\boldsymbol{ESA}(\boldsymbol{o}_i), \boldsymbol{ESA}(\boldsymbol{o}_j))$, to be used for fast similarity access. When the set S_n changes, as elements have been added or removed, M_s is updated. The similarity matrix allows to scale semantic content based recommendations up to millions of users.

USER EVALUATION

The afore described service prototype makes the linear radio experience more flexible, allowing the listener to modify timing, schedule and type of content, directly or indirectly. The user evaluation was limited to a subset of features, namely user interface elements, semantic content-based recommendations and proactive recommendations. The evaluation was divided into two parts: the first one involved a test of the prototype app and a survey, while during the second one we delivered a simplified version of the prototype, called “MyRadio”. In the following paragraphs we give more detail on the experiences.

First part: demo and survey

After a short introduction, the prototype app was demonstrated and tested by the listener. The researchers were available to answer to questions and to guide the test. After the test, each listener completed a short survey with questions about herself, her media consumption habits and the experience. Most of the questions were on a semantic differential Likert scale, allowing to reduce positive acquiescence bias and make questions easier to understand than using the traditional Likert scale [17].

The panel. 205 participants were involved: 11 groups during 14 months, mostly young students, the average age was 20 years with a standard deviation of 7 years. The young average age of the panel was reflected by their media consumption habits: only 32.2% listened to traditional radio while almost all listened to music. About 88.3% used YouTube to listen to music, 62% used Spotify at least one time. Also, 35% used a player app to listen to music. Only 12.2% listened to smartphone FM radio.

Figure 4 – The MyRadio trial app

Second part: extended trial

For the trial we developed a simplified version of the prototype app, the “MyRadio” app. The trial duration was about two weeks. Each listener interaction was recorded for the analysis. After the trial, a semantic differential Likert scale based questionnaire was given to the participants.

Proactive recommendations. MyRadio provided listeners with a mixed linear-AOD experience: each time the listener ran the app, there was one minute of linear radio on Rai RadioLive station, then the app automatically switched to a list of recommended AODs. The listener could skip the content at any time. Each listening session was about 8 minutes long: we tried to achieve an experience similar to watching some short movie trailers. Each recommended AOD was proposed for 90’’, after then a new AOD was presented. Recommended AODs were chosen among a list of about 2500 recent and automatically updated catch-up and archive audio content following a semantic content-based algorithm described in the following section.

Trial preparation. As the experience was based on recommended content, we tried to mitigate the cold start problem profiling the listeners before the trial. We used a personality based cross-domain correlation technique, see [18], [19]. Each listener answered to a preliminary questionnaire about her preferences on movies, books and music, and she was then included in a group of genre preferences out of five. The preparation reduced the cold start problem.

The panel. 50 participants were involved, equally divided between students from the University and Polytechnic of Torino and administrative and technical employees from Rai. The panel was randomly divided in two groups: a *test group*, with 74% of the participants, receiving content-based recommendations, and a *control group* with the remaining participants receiving random recommendations.

TRIAL MEASUREMENT RESULTS

Figure 5 includes skip and search, two of the most useful features the volunteers rated, compared with proactive recommendations. We clearly see the skip and search functions have been judged as the most appealing. In the full-text search functionality, the app also allowed to navigate the timeline of the AOD quickly reaching the keyword of interest, and 57% of volunteers found it very useful. Skip and proactive recommendations are heavily based on users' logs. Do listeners agree to leave their personal data? In the panel used for the tests, listeners generally agree to leave their listening history, more than personal demographic data, see Figure 6. Other feedbacks in the questionnaire also showed that volunteers were keen on getting useful functionalities for their personal data if required.

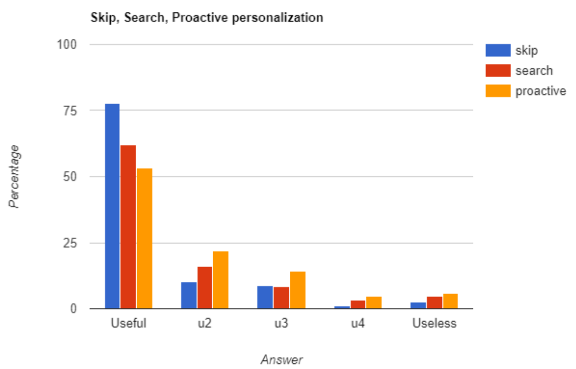


Figure 5 - Skip, search and proactive recommendations

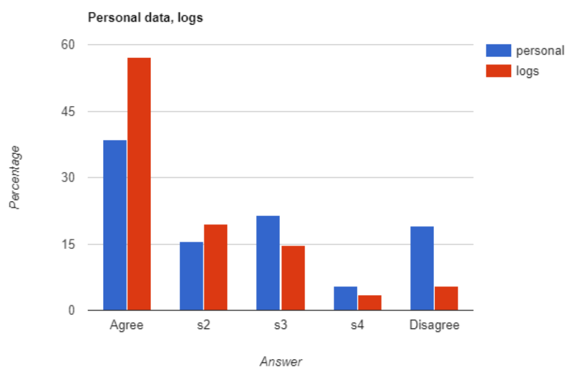


Figure 6 - Personal data and logs

Analyzing users' behaviour during the second phase of the tests (see Figure 7), we notice the control and test groups have different skip behaviours. The number of skips is nearly

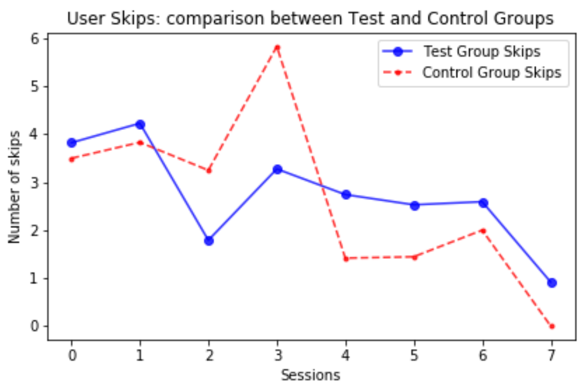


Figure 7 - Average number of skips per session (test vs. control group)

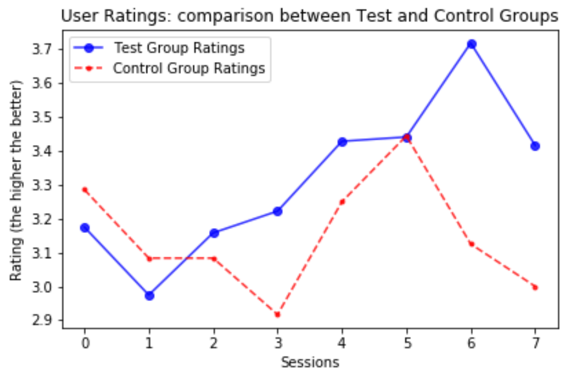


Figure 8 - Average listeners' ratings (test vs. control group)

the same during the first two sessions; then the number of skips increases in the control group. During the last sessions the test group shows a higher number of skips. This can be interpreted if we consider that the control group received random recommendations after the skip: after a high number of tries, they were giving up skipping. Moreover, semantic content-based recommendations have a clear effect on the rating: ratings are higher among the test group than in the control group, see Figure 8. In general, the possibility to skip content and receiving recommended AODs was very appreciated: 86% of the panel said it was useful or very useful. This can be related to the proposed recommended content: while the recommendation algorithm is not the most effective (content-based) nevertheless it increases listeners' satisfaction. The limits of the algorithm appeared when the genre relevance was compared with the content relevance: while the recommended clip genre was rated as useful by 48% of the panel, the content itself was rated as useful by 36% of the panel.

At the final survey, listeners were asked if the AODs relevance at the end of the trial was better than at the beginning: for 50% of them was slightly better, for 10% of them was better. Another interesting finding was related to AODs duration. About 50% of the panel was satisfied with the 90'' long AODs, while the main complaints were on content structure, e.g. too long preamble with programme jingle. Short audio clips were appreciated in both parts of the trial, as they could adapt to the available time and attention. Also, the most prominent limitation of current AODs emerged: *the lack of fine-grained metadata allowing to restructure the content depending on the context*. Labelled audio content atoms can be a solution, as well as thinking to content length adaptation from the beginning of content production. At the end of the trial, the test group rated a "service allowing to replace parts of the schedule with audio contents near to personal preferences" (proactive personalization) as very useful (28%) or useful (54%).

CONCLUSIONS AND FUTURE WORK

The paper describes a personalized, flexible radio service based on the HCR framework. After a short overview of radio personalization, the functionalities of the prototype were described. A constraint-based model describing context-aware proactive personalized radio was then discussed, showing a possible mathematical description and approximated solution. Finally, the results of a trial in two phases involving about 250 volunteers were discussed.

Some indications for future work emerged. First, audio content can be effectively adapted to the context if duration is flexible. Content divided in small slices with associated metadata, or *atomized* content, as well as content production with flexible duration in mind can be useful steps forward. Content with a more flexible structure seems to be a solution. Second, proactive context-aware personalization seems promising, however a more effective user experience has to be investigated: from more effective recommendations to an improved user interface. Finally, smart speakers with voice controlled digital assistants provide an interesting interface to deliver personalized radio, including proactively recommended content, and they will be investigated in the months to come.

BIBLIOGRAPHY

[1] P. Casagranda, A. Erk, S. O'Halpin, D. Born, W. Huijten, 2015. A framework for a context-based hybrid content radio. The Best of the IET and IBC Journal, 7, 41-47

[2] M. Armstrong, M. Brooks, A. Churnside, M. Evans, F. Melchior, M. Shotton. 2014. Object-based broadcasting – curation, responsiveness and user experience. Proceedings of 2014 International Broadcasting Convention

[3] EBU MIS, 2017. Audio-On-Demand, https://www.ebu.ch/publications/audio-on-demand (available for EBU members)

[4] ETSI, 2015. ETSI TS 102 818, Hybrid Digital Radio (DAB, DRM, RadioDNS); XML Specification for Service and Programme Information (SPI), ETSI

[5] R. Pagano, P.Cremonesi, M. Larson, B. Hidasi, D. Tikk, A. Karatzoglou, M. Quadrana. 2016. The contextual turn: From context-aware to context-driven recommender systems. In Proceedings of the 10th ACM conference on recommender systems, 249-252

[6] P. Casagranda, M. L. Sapino e K. S. Candan, 2017. Context Aware Proactive Personalization of Linear Audio Content, In Proceedings of the 20th International Conference on Extending Database Technology (EDBT), 574-577

[7] D. L. Strayer, J. Turrill, J. M. Cooper, J. R. Coleman, N. Medeiros-Ward, F. Biondi. 2015. Assessing cognitive distraction in the automobile. Human factors, 57(8), 1300-1324

[8] J. I. Arnason, J. Jepsen, A. Koudal, M.R. Schmidt, and S. Serafin, 2014. Volvo intelligent news: A context aware multi modal proactive recommender system for in-vehicle use. In Pervasive and Mobile Computing, 14. Special Issue on The Social Car: Socially-inspired Mechanisms for Future Mobility Services, 95-111

[9] F. Ricci, L. Rokach, B. Shapira, 2015. Recommender systems handbook. Springer

[10] C. P. Janssen, D. P. Brumby. 2009. Design and make aware: Virtues and limitations of designing for natural breakpoints in multitasking settings. In Accident Analysis and Prevention, 41, 115-122

[11] M. S. Young, K. A. Brookhuis, C. D. Wickens, P. A. Hancock. 2015. State of science: mental workload in ergonomics. In Ergonomics, 58(1), 1-17

[12] H. Kellerer, U. Pferschy, D. Pisinger. 2004. Knapsack problems. Springer.

[13] P. Casagranda, 2017. Hybrid Content Radio: A Framework for Personalized Linear Radio, PhD Dissertation, University of Turin

[14], S. R. Garzon. 2012. Intelligent in-car-infotainment systems: A contextual personalized approach. In Intelligent Environments (IE), 8th International Conference on, IEEE, 315-318

[15] E. Gabrilovich, S. Markovitch. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In Proceedings of the International Joint Conference on Artificial Intelligence, 7, 1606-1611

[16] D. S. Carvalho, C. Calli, A. Freitas, E. Curry. 2014. EasyESA: A Low-effort Infrastructure for Explicit Semantic Analysis. In Int. Semantic Web Conf., 177-180

[17] P. J. Danaher, V. Haddrell. 1996. A comparison of question scales used for measuring customer satisfaction. In International Journal of Service Industry Management, 7(4), 4-26

[18] P. J. Rentfrow, S. D. Gosling. 2003. The do re mi's of everyday life: The structure and personality correlates of music preferences. In Journal of personality and social psychology, 84(6), 1236

[19] I. Fernández-Tobías, M. Braunhofer, M. Elahi, F. Ricci, I. Cantador. 2016. Alleviating the new user problem in collaborative filtering by exploiting personality information. In User Modeling and User-Adapted Interaction, 26(2-3), 221–255

[20] D. Simon. 2013. Evolutionary optimization algorithms. John Wiley & Son

[21] Y. He, X. Zhang, W. Li, X. Li, W. Wu, and S. Gao. 2016. Algorithms for randomized time-varying knapsack problems. Journal of Combinatorial Optimization, 31(1):95-117.

ACKNOWLEDGEMENTS

The constrain based model has been discussed with Maria Luisa Sapino of the University of Torino and K. Selcuk Candan of Arizona State University. The trial with MyRadio has been carried out in collaboration with Michela Bellei during her internship in Rai.