

Personalization of content on public displays driven by the recognition of group context

Ekaterina Kurdyukova, Stephan Hammer, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Kurdyukova, Ekaterina, Stephan Hammer, and Elisabeth André. 2012. "Personalization of content on public displays driven by the recognition of group context." In *Ambient intelligence: Third International Joint Conference, AmI 2012, Pisa, Italy, November 13-15, 2012*, edited by Fabio Paternò, Boris de Ruyter, Panos Markopoulos, Carmen Santoro, Evert van Loenen, and Kris Luyten, 272-87. Berlin [u.a.]: Springer. https://doi.org/10.1007/978-3-642-34898-3_18.



Personalization of Content on Public Displays Driven by the Recognition of Group Context

Ekaterina Kurdyukova, Stephan Hammer, and Elisabeth André

University of Augsburg, Universitaetsstrasse 6a,
86159 Augsburg, Germany
{kurdyukova, hammer, andre}@hcm-lab.de

Abstract. Personalization of content on public displays relies on the knowledge of spectator interests and real-time recognition of social context. In busy public places, with numerous individuals circulating daily, the knowledge of individual interests becomes unrealistic. This paper presents an approach for automatic personalization which, instead of individual profiles, relies on *group context*. The system recognizes the constellation of spectators in front of a public display, based on their disposition and gender. Thus, the approach provides an important prerequisite for a completely automated personalization, requiring no input from the spectator side, neither for training, nor for real-time content adaptation. The experiment conducted in a public area showed that the presented approach can successfully identify the differences in the content observation of various groups. Moreover, the approach provides an insight into the diversity of circulating groups, and gives a hint about spectators' emotional and conversational response to the content.

1 Motivation

Personalized content on public displays offers clear advantages: the users get direct access to the information of their interest. A challenging task, however, is to *learn the interests* of the users and to *offer matching content* in real time. These tasks become even more complex if the displays are installed in busy public places, where numerous individuals circulate every day. Such displays have to learn the interests of a huge amount of users; moreover, they need to combine the interests of distinct individuals, when a group of several individuals observes the content.

Another challenge relates to the implicit nature of the personalization mechanism: the content adaptation should happen automatically, without any input from the user side. Indeed, a manual input, such as activation of an online profile or switching on a mobile client, is hardly acceptable in a busy public place. People have no time and attention for the manual input; moreover, they may be just unaware of the input possibility.

Therefore, there is a need in a mechanism that would *learn the interests* of the spectators and *adapt the content* completely *automatically*. Such a mechanism should distinguish between *different profiles* of spectators, however, requiring *no explicit input* from spectator side.

The system proposed in this paper meets the described requirements. In order to illustrate the concept, imagine the following scenario. A busy train station is equipped with a large public display. The display provides incoming passengers with city-related advertisement and useful tips. When a mother with two kids passes by the display, the screen advertises leisure activities for families. When a single lady passes by, the display shows a trendy perfume shop. For a passing-by couple, the display shows tips on romantic cafés. Finally, for a group of teenage boys, the display advertises an adventure attractions park.

The learning of spectators' interests and the real-time adaptation are enabled by the recognition of *group context*. By means of a camera mounted on the display, the system scans the composition of the observing groups: the number of individuals, their mutual disposition, and gender. Additionally, the system registers positive emotions of the individuals and whether they have a conversation.

In the learning phase the system tracks visual attention of the groups and relates it to the popularity of the content categories. During the real-time adaptation, the system recognizes the group standing in front of the display and shows the content which has the highest popularity within the given group.

In the current work we focus on the *learning phase*. By means of an experiment we demonstrate how the system can be used to identify differences in the visual attention (interest) of different groups. Important to mention, the notion of interest in this work equals to a spectator's "visual interest". We consider a person interested in the content, if he/she spends some time observing the content. Such an approach eliminates the confusion with person's intrinsic interest or commercial interest. Although the question of a person's real interest versus "visual interest" is definitely an important point, it is out of the scope of this work. The methods to measure audience's interest can be found in the related literature [1]. An argument against the "visual interest" can be the necessity of people to stand in front of the display, for instance, while waiting. The frontal orientation of the observing face, however, does imply the visual interest of the observer who keeps his or her face oriented to the content.

After an overview of related research, we describe the personalization mechanism in detail. In order to prove the usefulness of the system, we present the experimental deployment conducted in a public area of a university. Although this public space is definitely less busy than, for example, a train station, it does represent a valid public space with active and irregular circulation of diverse individuals. Therefore, this space is suitable to prove the performance of our system. The results of the experimental deployment show that the system can be successfully used to tag the content according to the group-based observation patterns (visual interest). Moreover, it provides interesting insights into the diversity of circulating groups and their emotional and conversational activity while observing the presented content.

2 Personalization of Content

In order to provide personalized content in real-time, a system should (1) learn the interests of the spectators and (2) be able to present the right content in the real-time according to the learnt interests. The first task is usually achieved by *tagging*. The second task refers to *real-time adaptation*.

2.1 Tagging

By means of tagging, the content elements are labeled with the spectators' interests. The interests can be retrieved in an explicit or implicit way.

For *explicit tagging*, a sample of potential users is asked to rank the presented content manually [2]. Such a tagging is usually done in laboratory conditions. The users rate each element of the content according to a given schema, e.g. a linear scale. Although the method delivers precise results, these results might not reflect the user interests. Since the lab setting is not natural, the ranking may deviate from the preferences the users would express in a real setting. Moreover, explicit ranking requires a significant effort from the user. The tiredness caused by the ranking routine, therefore, may also impact the number of provided ratings and the reliability of the results [2, 3].

Explicit tagging is barely applicable for the content of displays installed in a public place. The ranking reflects the interests only of some distinct individuals. Therefore, it can significantly deviate from the interests of the numerous other individuals circulating in the crowd. Another disadvantage refers to the extraction of group interests. Complex algorithms must be applied to derive the group interests from the interests of distinct individuals [4]. The complexity grows in busy public places where the group compositions are usually very diverse.

Implicit tagging is a more adequate ranking approach for public places. The method usually exploits crowd monitoring. An illustration of implicit tagging can be found in the work by Müller and colleagues [5]. The method counts the number of frontal faces registered for every content element (e.g. a slide). The element which has accumulated the largest number of the faces is considered to be of the highest interest. In the subsequent real-time adaptation process, the “most interesting” element will be set to the highest priority in the content schedule.

The approach replicates the real behaviour of the spectators: people look at the content when they are interested in it. However, the approach is not flexible enough to distinguish between groups of spectators. What if one content element was observed only by numerous single persons, and never – by couples? The approach, however, will prioritize the content element for both groups: singles and couples.

Another disadvantage of the method is the assumption that the frontal look equals interest. Although visual attention is an important hint to derive interest, it is not sufficient. In fact, the spectators can look at the display for many other reasons [6, 7]. For instance, the display is oriented frontally to the spectators' path or the colours of the content subconsciously attract attention. Such effects do not necessarily imply interest. Therefore, for automatic personalization more contextual cues should be used to support the assumption of interest.

To summarize, implicit tagging is the most suitable approach for public spaces. Such tagging requires no user input and reflects the natural setting. The tagging mechanism must distinguish the compositions of various spectator groups.

2.2 Real-Time Adaptation

The second part of the personalization process refers to the real-time content adaptation. The adaptation is based on the results obtained by the tagging. The real-time adaptation can rely on user *contributions* or work completely *automatically*.

The *contributions*-based approach requires a certain registration from the user side. For instance, the spectators switch on their Bluetooth devices and transmit the pre-set profiles [8, 9, 10]. The display receives the profiles and adjusts the content according to the profile interests. By means of dedicated strategies [11] the group interests can be derived from the individual profiles [12]. Although the contributions-based approach provides a precise overview about the present spectators, it is hardly applicable in a crowded public place. The numerous visitors of the public place may not possess the required devices. And even those who possess them may simply forget to switch the device on. As a result, the display will retrieve an incomplete or a wrong picture about the surrounding spectators.

The alternative *automatic* approach can utilize the identity of the user. For example, by means of face recognition the system can understand who stays at the display, and thus automatically adjust the content [13]. Although such an approach does not require any user contribution, it has to carefully learn the user profiles in advance. This requirement is not realistic in a busy public place with numerous individuals. Müller and colleagues [5] used face detection as a trigger for real-time adaptation. Once a face is detected in front of the display, the most popular content appears on the screen. The method is more suitable for public places than identification: it eliminates the unrealistic knowledge of each single individual. However, it does not distinguish between the interests of various individuals. For instance, the content popular among women does not necessarily match the content popular among men. Moreover, the method cannot take into account the composition of the spectator groups. For example, the same content will be displayed for two teenage girls, a couple, a group of elderly men or a mother with three kids.

All in all, it is a challenging task to recognize the interests of spectators and provide the right content in the real-time. Ideally, the system should know the individual profiles of the spectators, which is difficult to realize in a busy public place. Moreover, the system should work automatically, not requiring any input from the spectator side. A trade-off would be an approach which approximates to the individual profiles, but does not require the spectators to explicitly provide the system with their profiles.

In this paper we present a group-based personalization approach, focusing on the tagging phase. Based on the gender and disposition of detected spectators, the system classifies the spectators into distinct groups, for instance, a couple, two men or a single woman. Thus, the system registers the interest not of distinct individuals, but of distinct groups. Since the detection of the spectators and the recognition of their gender are done with a camera, the approach works fully automatically eliminating any user input.

3 Approach for Group-Based Personalization

Based on the analysis of the existing approaches for tagging and real-time adaptation, we came up with a set of requirements for personalization systems in busy public places:

- the system must work **completely automatically**, not expecting any contribution from the user (spectator) side.
- the system must **capture the group composition**, instead of identities of distinct individuals. The unique groups should be defined by the characteristic features of the group members, e.g. gender or age.
- the system must be **robust against the traffic of a busy public place**, taking into account the diversity of passing-by individuals and possible groups.

The system presented in this paper meets the requirements of automatic personalization. Below we describe how it can be used for automatic tagging and real-time adaptation.

Completely Automatic System

Content tagging and subsequent adaptation are based on the monitoring of spectators. A camera installed on the top of the display detects the faces in the proximity of the display (see Fig.1); the microphone detects the noise level.

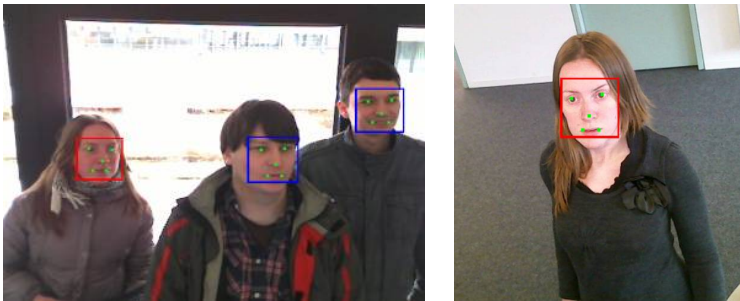


Fig. 1. Recognition of the group structure: red and blue rectangular indicate female or male faces

The system runs face detection algorithm enabled by the SHORE [14] and SSI [15] software. SHORE provides robust face detection and gender recognition. For each detected face, it delivers the following parameters:

Gender. The SHORE software delivers a probability percentage for the face being male or female. Our system uses the threshold of 80% to accept SHORE's decision on gender.

Position. The SHORE software provides coordinates of outlining rectangle for each detected face. From the size of the rectangle and its position on the x-axis we can infer the spectators' proximity to the display, location, and mutual disposition.

Emotional state. The SHORE software also provides emotion recognition, relying on facial expressions¹. The software delivers percentage on probability for four emotions: “happy”, “angry”, “surprised”, and “sad”. If the probability is lower than a pre-defined threshold, the emotional state is registered as “neutral”. If the probability of an emotion is higher than the set threshold (we used 70%), the face is registered as expressing the given emotion.

Besides information on detected faces, the tagging system registers whether spectators have a *conversation*. A microphone integrated into the camera registers the volume level when spectators are present in front of the screen. The volume level is classified into three ranges: silent (almost no sound), moderate (moderate discussion between several individuals), and loud (active discussion).

The SSI software helps synchronize the content slide show with the data delivered by SHORE. The SHORE and SSI software run in the background, leaving in the foreground solely the content. Thus, the system runs the group recognition in a complete automatic way. It neither requires any input from the spectators, nor does it disturb the observation process.

Capturing the Group Composition

Having the data on present faces, their gender, and disposition, the system can conclude about the group composition. The spatial disposition of the faces enables us to determine whether the present spectators belong to the same group or are standing alone individuals.

Robust against the Traffic of a Public Place

Generally, the number of SHORE-detected faces is not limited. The software, however, allows defining a minimal size of the face outline as a percent of the entire camera field. We set the smallest face to be 2,5% of the field covered by the camera. Smaller faces refer to distant persons who cannot see the content properly; thus, they are not considered as valid spectators.

The recognition of spectator’s interest can be supported by additional cues, such as positive emotions or discussion of the content. Our personalization mechanism registers emotions of each group member, as well as the volume level of the conversation. However, the reliability of these additional cues has to be proven experimentally. Emotions and conversations are not necessarily caused by the display content. Therefore, we consider emotional and conversational response as a secondary hint to the spectator’s interest.

Speaking about emotional response, it is important to mention that the interest does not always imply positive emotions. For instance, a person can be highly concentrated on the content (thus, interested), but have a neutral facial expression. A positive emotion therefore is not equivalent to “relevant”, but is a contextual condition that influences how “relevant” the content is.

¹ The details on implementation are given by the SHORE authors in the related literature [14].

4 Accuracy of Recognition

Before the deployment of the system, we tested the accuracy of face detection and emotion recognition delivered by the SHORE. Important to emphasize, our goal was not to verify the accuracy of the SHORE algorithms. This question has been elaborated by the authors of SHORE and can be found in the related literature [14]. Our goal was solely to test how accurate the SHORE software performs in our experimental conditions.

The test was conducted in the public area of the university, where the main experiment took place, employing the same displays as in the main experiment. For the test we presented some arbitrary photos on the displays. All spectators were video-recorded; a note informed them about the recording fact. Simultaneously with the video recording, the tagging system was running in the background. It logged the detected faces, their gender, and emotions.

In total, we collected 16 hours of video, containing 128 female and 120 male faces. The video material was manually annotated, registering the recorded number of female and male faces, group constellation, and emotions (based on subjectively estimated facial expressions). The annotation was compared with the log data, yielding the accuracy of recognition.

About 95% of all the faces of people standing more than 1.5 seconds in front of the display were captured by SHORE. The faces further than the specified observation distance (face rectangle covering less than 2.5% of the camera view) as well as the faces of passers-by who just glanced at the display without stopping were not detected by SHORE as faces. This limitation is in line with our definition of spectators: people within a close proximity to the screen, who do stop to watch the content.

Gender Recognition of SHORE showed accuracy rate of 96% for males, and 92% for females. The system needed about 0.3 seconds on average to decide on the gender.

Emotion Recognition of SHORE, i.e. for “happy” emotion, showed an accuracy rate of 90% for male spectators and 92% for female spectators. The recognition of “surprised”, “sad”, and “angry” emotions showed less reliable results, yielding only 60-65% of accuracy. Moreover, these emotions were recognized on the video rarely: usually, spectators either smiled or didn’t express any emotions.

To summarize, gender recognition with SHORE can be reliably used for the group-based personalization. As for the emotion recognition with SHORE, we can reliably apply only the recognition of the classes “happy” vs. “not happy” emotion (in other words, “smile” vs. “no-smile”).

5 Experimental Deployment

The goal of the experimental deployment was to see how well the system can be applied for group-based personalization. In particular, we aimed to answer the following questions:

- Can the system identify differences in observation patterns (content preferences) of distinct spectator groups?
- What insights into spectator groups can be gained with the system?
- Does information on emotional and conversational response deliver reliable hints on spectator interests?

5.1 Deployment Set-Up

The system was deployed during three weeks on the displays in a university public area. Three displays were involved in the deployment: one display situated in a lobby and two in a passage (see Fig. 2). All displays have non-touchable screens of 62 inches and 45 inches in diagonal.



Fig. 2. The lobby display (left) and the passage displays (right) used in the experiment

The circulation of people on the premises of the university is moderate. Besides the main “inhabitants”, consisting of about 30 researchers, the experiment area is used by students and visitors. The passage area is often used as a short cut to the university canteen, the parking lot or other places within the university. During the experimental weeks two events took place at the area adjacent to the experiment public place; bringing in total about 30 visitors from outside the university.

The aim of the personalization system was to tag the content newly created for the university displays. The content was compiled in a slide show; the personalization system ran on the background. The content topics were proposed by researchers of the university. Within a brainstorming session, the researchers came up with four content categories: “Team”, presenting the members of the research team, “News”, informing about recent info, e.g. upcoming events or lectures, “Department Life”, presenting events of research unit, “Quiz”, posting a tricky question about a research unit, followed by the correct answer. The researchers found these categories relevant for the university life. However, we needed to find out whether the content would also attract our students and visitors.

The design of the content was kept consistent, in order to exclude distractions caused by visual design (see Fig. 3). Each content slide stayed on the screen for 10 seconds.



Fig. 3. Examples of the content: “Department Life” and “Quiz”

5.2 Tagging Procedure

In order to tag the content categories according to the group interests, the tagging algorithm was launched in the background of the slide show. Each display was supplied by a camera with an integrated microphone. The cameras were installed on the top side of the display frame. The SHORE software was processing the images captured by the camera. With the frequency of 15 frames per second, SHORE delivered information on each detected face: gender estimation, coordinates of outlining rectangular, and emotion estimation.

This data was processed to make an entry to the log file. Based on the number of detected faces and the gender data, the **<group composition>** was calculated. Based on the coordinates of the rectangular, we calculated **<position>** of each group member. Position reflected the user location at the display (left, centre, right) and the proximity to the display (near, middle, far). From this information we could estimate whether the spectators belong to the same group (stand next to each other) or are several distinct individuals. Based on the probability of each emotion, we registered the resulting **<emotion>**. The emotion having probability more than 70% was entered to the log. Finally, the microphone provided data on estimated **<volume level>**.

As a result, a log entry consisted of the general description of the social context and the detailed description of each face:

```
<timestamp> <group composition> <volume level>
  <face1><gender><position><emotion>
  <face2><gender><position><emotion>
```

An entry was added to the log every time the social context was changing, for instance, people joined the group, people left, or emotional context changed. The following lines illustrate an example of a log entry (F stands for female, M stands for male):

```
<15:29:00> <2F + 1M> <loud>
  <face1><F><left near><neutral>
  <face2><F><left near><neutral>
  <face3><M><left near><happy>
```

The log files were created for every day, separately for each display. The tagging system did not capture any raw video and audio signals.

After the experiment, the log files were parsed. We summarized how frequently the groups observed each content category, which emotions were expressed, and the volume of the conversations. Additional information, such as the number of all groups, the total number of females, etc. could also be derived from the log files.

6 Experiment Results

In total, 324.2 hours and 4727 detected faces were recorded in the log files. The analysis of the log files enabled us to answer all questions posted to the experiment. First, we proved that the system is able to recognize the interests of distinct spectator groups. Second, we obtained interesting insights for the groups circulating in the public area. Finally, we could conclude whether emotional and conversational context can support the evidence of spectators' interests. Below we provide the detailed results.

The system successfully identified the differences in observation patterns (visual interests) among distinct groups. Figure 4 illustrates the distribution of visual interests among groups of two or three spectators. The illustration clearly shows the differences in observation patterns: for instance, topic "News" was more frequently observed by the group "1 Male + 1 Female" than by other groups. Topic "Quiz" was more often observed by homogeneous groups, "2 Males" or "2 Females". The distribution refers to the data obtained at the lobby display; very similar distribution patterns were observed on the passage displays. The figure reflects the interests of only composite groups; the interests of single spectators (one male or one female) were distributed similarly to the interests of the respective groups of two (two males or two females).

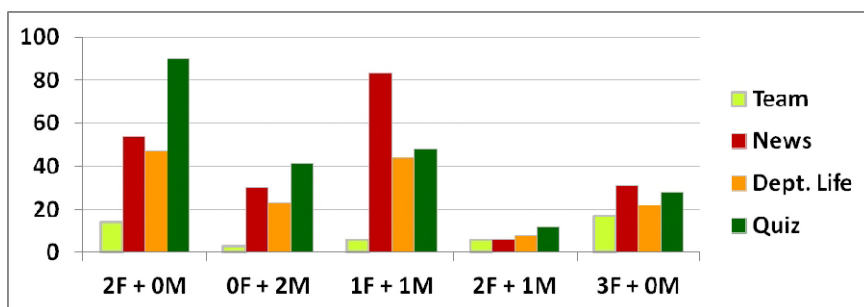


Fig. 4. Distribution of group interests (F stands for female, M – for male). Y axis indicates the number of times the visual interest of the group was detected

In total, the system detected 10 different kinds of groups. The majority of the detected spectators were single individuals (see Fig. 5). This finding was quite surprising for us, since many meetings and collaborations take place at the university.

Observing the behaviour of spectators, we realized that in spite of the gathering in meeting rooms or lecture halls, the transitional public places (such as the passage and the lobby) people mostly pass alone. Detected composite groups consisted mostly of two persons.

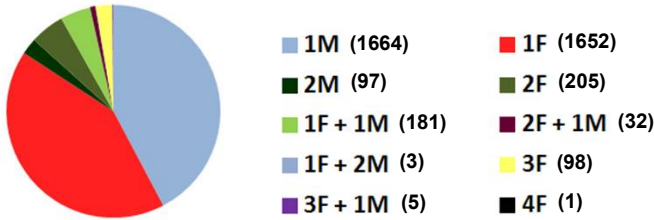


Fig. 5. Distribution of spectator groups. M stands for male, F - for female

During the experiment the system registered a solid number of positive emotions. For the analysis, we considered only positive emotions, since the SHORE software yields reliable recognition results only for “happy” vs. “not happy” emotions (see Section 4).

Figure 6 illustrates the distribution of positive emotions among single female and single male spectators. We conducted the analysis on single males and females, since they were the most represented spectator types. The analysis for other groups can be done similarly.

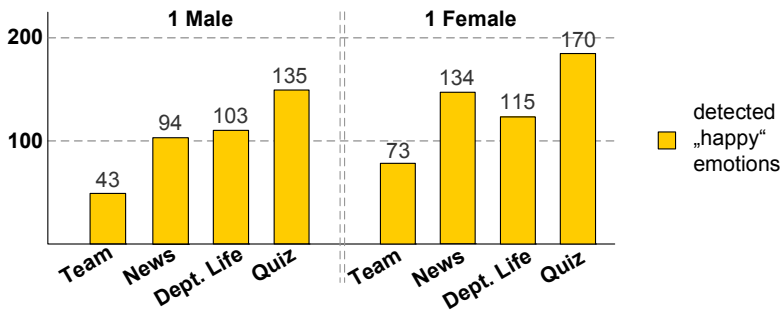


Fig. 6. Distribution of “happy” emotions

From the first sight, Figure 6 uncovers clear differences in the frequency of positive emotions expressed for different content topics. However, calculating conditional probabilities for each topic (considering how frequently each topic was observed by either group) we didn’t find any noticeable differences.

Male spectators showed almost equal emotional response to all content topics: “Team” (0.34), “News” (0.34), “Department Life” (0.36), “Quiz” (0.33). Females had a lightly more frequent positive response to “Team” (0.53) and “News” (0.45); however, quite a similar response to “Department Life” (0.4) and “Quiz” (0.39). Generally, we found that males expressed positive emotions slightly less frequently than females (0.34 and 0.43).

The analysis of the conversational activity was done in a similar way. For the analysis we considered the groups of two spectators. We chose these groups for the analysis, since the system mostly detected conversations between two persons. For each group we calculated conditional probabilities: how often and in which volume a group had conversations while observing the content.

The analysis did not reveal any noticeable differences. Most of the conversations were done in moderate volume, independently on the content topic.

Homogeneous groups (only males or only females) were slightly more silent when observing the content “Quiz”. Mixed groups had generally slightly more conversations when observing “News”. These observations can be explained by the nature of the content. “Quiz” posts the spectator a question, substituting a real conversation and thus making people silent. “News” provokes a discussion about some urgent events. However, the conversations could also be not related to the content.

7 Discussion

Below we provide the interpretation of the experiment results, addressing the research questions posted above. We discuss limitations of the study, further steps, and possible applications of the presented approach.

7.1 Content Preferences of Distinct Spectator Groups

The experiment has shown that the group-based personalization mechanism can successfully extract the differences in observation patterns of distinct spectator groups.

The main interest differences can be observed between homogeneous groups (only males or only females) and mixed groups (a male and a female). Homogeneous groups mostly preferred “Quiz” category, whereas the mixed groups were more interested in “News”.

The phenomenon can be explained by the relationships within homogeneous and mixed groups. Observing our spectators, we noticed that homogeneous groups often represent close friends. They meet at the university not only for study-related occasions, but also for socializing, chatting or spending a free time slot together. Therefore, they are likely to involve into such an entertaining occasion as a quiz. Mixed groups often represent study fellows, connected not by a friendship, but rather by a common studying activity. They meet at the university for a certain study-related occasion, e.g. to work on a project or prepare for an exam. Therefore, they are not likely to spend time for a “Quiz”, but would rather pay attention to the study-related “News”.

The overall majority of spectators showed more interest to the content “Quiz” and “News”. The preference to “News” relates to its informative content: people tried not to miss relevant and important facts. The preferences to “Quiz” can be explained by its interactive nature. The quiz questions were related to the university stories. Therefore, the quizzes not only challenged the spectators, but also gave them some curious facts.

Comparing these results with the observations of Rist and colleagues [16], who evaluated various contents at university displays, we may see slight contradictions. The authors reported that people *generally* have lower interest in entertainment content and higher interest in news. However, in their work, entertainment related to the games which demonstratively uncover user participation. Unlike games, our entertaining quiz allows the users to participate unnoticeably, with no demonstration of success or failure. Such unnoticeable interaction is known to be appreciated by people in public locations [17, 18].

7.2 Gaining Insight into Spectator Groups

Detected groups contained slightly more females than males. We found this fact surprising: statistically our technical institute counts more males than females. One explanation of this phenomenon can be the natural curiosity of women and their ability to notice the surrounding objects better than men [19].

Analyzing the log files we could see that single spectators were often joined by other persons, creating a group. Such behavior is known as the “honey pot effect” [20, 21]. People are not courageous enough to demonstrate their interest in public. Thus, they feel more comfortable to join an existing spectator.

Among the 10 detected groups, only 5 groups were presented in the passage area. The circulation of people in the lobby is indeed higher, since it is a large recreation room where people usually gather. The passage, on the contrary, is a narrow corridor. People usually pass it quickly, heading to a certain room or to the canteen. The lower number of spectators in the passage can also be explained by the orientation of the displays. As mentioned by Müller et al. [22], the displays oriented at 180 degrees to the user trajectory attract less attention than the displays oriented at 90 degrees. This observation applies to the orientation of our displays: the passage displays are oriented at 180 degrees, and the lobby display – at 90 degrees to a typical passer-by trajectory.

7.3 Emotional and Conversational Response

Observing arbitrary spectators, we noticed that positive emotions and conversations are often not related to the content. They are usually brought from a dialogue preceding the display observation. Therefore, our experiment results do not give enough evidence that detected positive emotions and conversations were *provoked* by the content.

7.4 Application in Other Public Spaces

Although the experiment was conducted in a public space with rather moderate circulation of people, it demonstrates that the system can be deployed in other public spaces. Apart from the university public space scenario, the system can be applied in an environment with a brighter diversity of groups.

The system installed at a large *shopping mall* can recognize the interests of different customer groups. Unlike existing ambient technologies facilitating shopping experience [23], our system is able to learn the interests of the customers. Based on the

learnt shopping interests, the system can advertise the matching content immediately when customers approach the display. In a similar way, the system can be deployed at a *travel agency*. It will help to recognize trends in vacation destinations among couples, single travelers, families, etc.

The system can give an insight into the tastes of the people. Imagine the system installed at a *picture gallery* or a photo exhibition. Tracking how visitors observe the art pieces, we can conclude which authors and which genres are popular among different visitor groups. Such information could facilitate planning of the future exhibitions.

Finally, the system has a potential to impact the tastes of the people. Imagine the system to be installed at a university “*Open Doors*” day. The “Open Doors” day is an annual event organized by universities, aimed to orient school students in the choice of their future education. A current problem of engineering faculties is a low ratio of female students. The problem is partially caused by gender stereotypes, but partially by insufficient awareness of school girls about the engineering career. A display recognizing social context could increase their awareness. Once girls are recognized in front of the screen, the display can switch to the Engineering content.

7.5 Real-time Adaptation

The experiment illustrated how the system can be used for content tagging. The next step, real-time adaptation, can be achieved by integrating the extracted preferences into the adaptive content schedule. Once a group approaches the display, the system recognizes the group structure and switches to the content preferred by the group.

In order to validate the system performance for real-time adaptation, a more realistic *public* setting is necessary. The experiment presented in this work does show that the group-based approach can be successfully applied in a public setting. However, we found that the groups presented at the university environment are not that diverse. A real busy public place, such as a train station or a shopping mall, would be more appropriate to test the system in real-time adaptation mode.

8 Conclusion

The paper presented a system for group-based personalization on public displays. The system can be used for the *tagging of content* according to spectators’ interests and for the real-time *content adaptation*.

The advantage of the proposed invention over existing systems is its completely *automatic* adaptation mechanism. The extraction of interests, as well as the real-time adaptation is performed automatically, without any input from spectator side. This requirement is critical in busy public places: the passers-by are unlikely to have time, attention or means for an input.

Another advantage of the proposed system is its capability to distinguish between spectator profiles. Instead of the retrieval of individual profiles (which is hardly realistic in a busy public place), the system extracts *groups profiles*. The groups are defined according to the number of spectators, their disposition, and gender. For example, the display distinguishes between two women, a couple, a girl or a group of boys.

An experimental deployment, conducted in a real public space, proved that the system can successfully identify the differences in observation patterns (visual interests) of different spectator groups. Moreover, the experimental results gave us insights into the circulating groups: what constellations of spectators are typical, which groups circulate in different public areas, what is the proportion of female and male spectators.

Finally, the experiment enabled us to conclude whether the tagged data on the emotional and conversational response can be correlated with the displayed content. The results gave us no evidence that positive emotions and conversations are directly related to the content. Often they are caused by events preceding the display observation. However, we have shown that the system can reliably tag positive emotions. We believe that a more entertainment-oriented content (such as a photo exhibition) can reveal differences in emotional response. The content chosen for the experiment was rather emotionally neutral; it addressed the topics relevant to the public area – a university environment.

As the next step we are planning to extend the definition of groups by recognition of *age*. This advance will enable us to distinguish, for example, between two teenage boys, an old couple or a mother with a kid. These groups are likely to have very different content preferences.

Moreover, we are planning to run a study on a real-time content adaptation, in order to see how spectators accept the automatic adaptation. However, the study has to be run in a more real public place, with diverse group profiles.

Acknowledgements. This research is partly sponsored by OC-Trust project (FOR 1085) of the German research foundation (DFG).

References

1. Müller, J., Alt, F., Michelis, D.: *Pervasive Advertising*. Springer (2011)
2. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: *Recommender Systems: An Introduction*. Cambridge University Press (2010)
3. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., Riedl, J.: Getting to know you: learning new user preferences in recommender systems. In: 7th International Conference on Intelligent User Interfaces, pp. 127–134. ACM Press, New York (2002)
4. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *User Modeling and User-Adapted Interaction* 14(1), 37–85 (2004)
5. Müller, J., Exeler, J., Buzeck, M., Krüger, A.: ReflectiveSigns: Digital Signs That Adapt to Audience Attention. In: Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B., Tobe, Y. (eds.) *Pervasive 2009*. LNCS, vol. 5538, pp. 17–24. Springer, Heidelberg (2009)
6. Huang, E., Koster, A., Borchers, J.: Overcoming Assumptions and Uncovering Practices: When Does the Public Really Look at Public Displays? In: Indulska, J., Patterson, D.J., Rodden, T., Ott, M. (eds.) *PERVASIVE 2008*. LNCS, vol. 5013, pp. 228–243. Springer, Heidelberg (2008)
7. Müller, J., Alt, F., Michelis, D., Schmidt, A.: Requirements and design space for interactive public displays. In: 9th International Conference on Multimedia, pp. 1285–1294. ACM Press, New York (2010)

8. Alt, F., Balz, M., Kristes, S., Shirazi, A.S., Mennenöh, J., Schmidt, A., Schröder, H., Goedicke, M.: Adaptive User Profiles in Pervasive Advertising Environments. In: Tscheligi, M., de Ruyter, B., Markopoulos, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (eds.) *AmI 2009*. LNCS, vol. 5859, pp. 276–286. Springer, Heidelberg (2009)
9. Karam, M., Payne, T., David, E.: Evaluating BluScreen: Usability for Intelligent Pervasive Displays. In: 2nd International Conference on Pervasive Computing and Applications, pp. 18–23. IEEE Press, New York (2007)
10. Mahato, H., Kern, D., Holleis, P., Schmidt, A.: Implicit Personalization of Public Environments using Bluetooth. In: 26th Annual Conference on Human Factors in Computing Systems, pp. 3093–3098. ACM, New York (2008)
11. Jameson, A., Smyth, B.: Recommendation to Groups. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 596–627. Springer, Heidelberg (2007)
12. Villar, N., Kortuem, G., Van Laerhoven, K., Schmidt, A.: The Pendle: A Wearable Mediator for Mixed Initiative Environments. In: International Workshop on Intelligent Environments, pp. 173–181. IEEE Press, New York (2005)
13. Vogel, D., Balakrishnan, R.: Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: 17th Annual ACM Symposium on User Interface Software and Technology, pp. 137–146. ACM, New York (2004)
14. Küblbeck, C., Ernst, A.: Face detection and tracking in video sequences using the modified census transformation. *Image and Vision Computing* 24(6), 564–572 (2006)
15. Wagner, J., Lingenfeller, F., André, E.: The Social Signal Interpretation Framework (SSI) for Real Time Signal Processing and Recognition. In: 13th Annual Conference of the International Speech Communication Association, pp. 3245–3248. Springer, Heidelberg (2011)
16. John, L., Rist, T.: xioScreen: Experiences Gained from Building a Series of Prototypes of Interactive Public Displays. In: *Ubiquitous Display Environments*. Springer (2012)
17. Ning, T., Müller, J., Walter, R., Bailly, G., Wacharamanotham, C., Borchers, J., Alt, F.: No Need To Stop: Menu Techniques for Passing by Public Displays. In: Workshop on Large Displays in Urban Life at International Conference on Human Factors in Computing Systems (2011)
18. Holleis, P., Rukzio, E., Otto, F., Schmidt, A.: Privacy and Curiosity in Mobile Interactions with Public Displays. In: Workshop on Mobile Spatial Interaction at International Conference on Human Factors in Computing Systems (2007)
19. Pease, A., Pease, B.: *The Definitive Book of Body Language: How to Read Others' Attitudes by Their Gestures*. Sheldon Press, London (1988)
20. Mathew, A., Rogers, Y., Lloyd, P.: Post-it note art: evaluating public creativity at a user generated art installation. In: 8th ACM Conference on Creativity and Cognition, pp. 61–70. ACM Press, New York (2011)
21. Müller, J., Walter, R., Bailly, G., Nischt, M., Alt, F.: Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In: International Conference on Human Factors in Computing Systems, pp. 297–306. ACM Press, New York (2012)
22. Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., Krüger, A.: Display Blindness: The Effect of Expectations on Attention towards Digital Signage. In: Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B., Tobe, Y. (eds.) *Pervasive 2009*. LNCS, vol. 5538, pp. 1–8. Springer, Heidelberg (2009)
23. Meschtscherjakov, A., Reitberger, W., Mirlacher, T., Huber, H., Tscheligi, M.: AmIQuin - An Ambient Mannequin for the Shopping Environment. In: Tscheligi, M., de Ruyter, B., Markopoulos, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (eds.) *AmI 2009*. LNCS, vol. 5859, pp. 206–214. Springer, Heidelberg (2009)