Designing Prototype Player Personas from a Game Preference Survey

Joni Salminen  
Qatar Computing Research Institute, Doha, Qatar; and University of Turku, Turku, Finland  
jsalminen@hbku.edu.qa

Soon-gyo Jung  
Qatar Computing Research Institute  
Doha, Qatar  
sjung@hbku.edu.qa

Jukka Vahlo  
University of Tampere  
Tampere, Finland  
jukka.vahlo@staff.uta.fi

Shammur A. Chowdhury  
Qatar Computing Research Institute  
Doha, Qatar  
shchowdhury@hbku.edu.qa

Aki Koponen  
University of Turku  
Turku, Finland  
aki.koponen@utu.fi

Bernard J. Jansen  
Qatar Computing Research Institute  
Doha, Qatar  
bjansen@hbku.edu.qa

Abstract  
The competitiveness of the video game market has increased the need for understanding players. We generate player personas from survey data of 15,402 players’ 195,158 stated game preferences from 130,495 game titles using the methodology of automatic persona generation. Our purpose is to demonstrate the potential of data-driven personas for segmenting players by their game preferences. The resulting prototype personas provide potential value for game marketing purposes, e.g., targeting gamers with social media advertising, although they can also be used for understanding demographic variation among various game preference patterns.

Author Keywords  
Video game; player personas; automatic persona generation

CSS Concepts  
- Human-centered computing~Human computer interaction (HCI)

Introduction  
Although personas are shown to be advantageous to many use cases in design, software development, and gaming [5, 21], their application is prohibited by time and high cost of creation. Therefore, many startups and
### Related Literature

**Data-driven personas** are based on empirical data of large quantities, often focusing on behavioral metrics and demographic attributes [2], addressing data limitation of qualitative persona creation [4]. Data-driven personas can be created from system logs and organizational records describing the users or customers [3]. For example, Molenaar [15] analyzed 400,000 clickstreams, grouping them into common workflows and classifying users into these workflows.

Using a quantitative approach, Zhang et al. [22] analyzed clickstream data with hierarchical clustering to identify common click flows and generate data-driven “personas” with names. However, their personas only include a manually given name but no other personified information, such as age, gender, location, and motives. This limitation also applies to procedural player personas that are generated e.g. with evolutionary algorithms and neural networks [10]. These personas are manually given names based on their behaviors of playing the game (e.g., “monster killers”). Rather than human beings with name and personified information, these personas are more like virtual agents that capture the behavioral variation within the game playing context [19]. Therefore, creating player personas with personified attributes is an open research question, with limited prior work.

**Player personas** are intended for stakeholders interested in gaming audiences, including game designers and developers, marketers, and players who wish to understand other players. Particularly in the gaming context, personas can be used to represent player archetypes involving player attributes and game choice preferences [20, 21]. In previous research,

---

**Table 1: Benefits of automatic persona generation**

<table>
<thead>
<tr>
<th>Manual personas</th>
<th>Automatic personas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low sample size (&quot;small data&quot;)</td>
<td>High sample size (&quot;big data&quot;)</td>
</tr>
<tr>
<td>Qualitative data</td>
<td>Quantitative data</td>
</tr>
<tr>
<td>Slow to create (typically taking months)</td>
<td>Fast to create (typically taking days)</td>
</tr>
<tr>
<td>Unresponsive to changes in user preferences</td>
<td>Responsive to changes in user preferences</td>
</tr>
<tr>
<td>Expensive</td>
<td>Affordable</td>
</tr>
</tbody>
</table>

| Total number of preferences  | 195,158                      |
| Number of unique respondents  | 15,402                       |
| Average number of preferences per respondent | 12.67                      |

**Table 2: Description of the dataset**

Small companies (e.g., independent game developers) cannot afford the investment of a full-scale persona project. Consequently, many small game development companies may not be able to leverage personas or they have to rely on personas based on ‘ad hoc’ intuitions rather than real player data.

Automatic persona generation (APG) aims to bring personas within the reach of smaller organizations such as indie developers [1, 2], while both enhancing customer-driven decision making and helping organizations achieve their goals in competitive online markets (see Table 1). APG is a methodology and system for creating personas from digital data [1, 2]. It is specifically developed to address the limitations of manual persona creation. Personas from APG are (1) rapidly created (in a matter of hours rather than months), (2) accurate, and (3) can be updated through fresh datasets or connecting to live APIs [11].

Our research goal is to apply APG to a new context to develop player personas. We also address the shortcoming of previous player persona research, namely the fact that the personas in previous research [10, 13] are missing personified attributes (e.g., picture, name and other human attributes). To address these gaps, we collect survey data on players’ game preferences (i.e., their favorite games) and apply the APG methodology to develop data-driven personas from this dataset. This approach has value for gaming companies that can use the generated personas to create more empathetic messages [14] and apply specific targeting criteria corresponding to personas [17] to increase the effectiveness of player acquisition in the competitive video game industry.
players personas have mostly been created using game analytics tools that typically focus on the effectiveness of the level design, player behavior, monetization, and game performance at large. Many companies combine the former with play testing, questionnaires, interviews, and psychometrics that are utilized across game development phases. For example, tuning the difficulty level of a game is known to be an important factor for successful game design, and companies use telemetric data and game log data to find an appropriate level of difficulty to fit the needs of their core audiences [6].

Drachen et al. [7] applied personas to identify player profiles based on player behavioral factors, e.g., completion time and number of deaths. The challenge for game behavior-based player personas is that it is difficult to extend the identified constructs to other games [6]. This challenge is largely because procedural player personas are typically constructed based on specific characteristics of a single game [10, 13] providing limited tools for understanding game preferences of particular player segments. To holistically understand player segments, the data should be based on players’ preferences to play a variety of games instead of their behaviors when playing a single game. Moreover, the purpose of procedural player personas tends to be simulating player behaviors [10, 13]. This purpose, while technically interesting, lends itself poorly for decision-making where players should be understood as people rather than abstract patterns (e.g., for video game marketing). Thus, novel methods to generate player personas are needed.

Methodology

Data Collection

Data was collected using an online form that enabled the respondents to choose their favorite games. The respondents also indicated their age, gender, and preferences in gameplay elements. The 15,402 respondents were recruited from Finland by marketing the online form as a player type test. By taking the test on www.kinrate.com, respondents were shown their player type and game recommendations based on similar players. Respondents selected their favorite games by typing into a form field, triggering a query to the underlying database of 130,495 game titles that were shown as autocomplete suggestions. The game titles were obtained from Internet Game Database (IGDB) using the API service (https://www.igdb.com/).

The final dataset includes 195,158 video game selections by the respondents (see Table 2).

Persona Generation

To generate player personas from the survey data, our methodology undertakes several steps (see Figure 1).

First, we create an interaction matrix where rows are players and columns are games chosen by each player. For example, if Respondent 01 chose "Halo" as their favorite game but Respondent 02 did not, the matrix would have the value of “1” for the cell <R01-Halo> and “0” for <R02-Halo>. Because we have p=15,402 players and g=130,495 possible games, the matrix has p x g = 2,009,883,990 cells.

Second, after creating the interaction matrix, we group the games by gender (male, female) and age group. The age groups (13-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+) are compatible with the typical age groups available in online analytics platforms [2].
Figure 2 illustrates the process of transforming individual game preferences into grouped instances.

After this grouping, we apply non-negative matrix factorization (NMF) [12] for identifying latent game preference factors that form the basis of the player personas (see Figure 3). Earlier studies show the applicability of NMF for persona generation [1, 2]. NMF reduces the dimensionality of large datasets by replicating the original data structure using latent factors. These latent factors represent the variation in the players’ game selections and can be interpreted as archetypes of game preferences. The NMF factors, therefore, form the basis of the personas. Note that because the NMF relies on game preference data grouped by demographic attributes, the resulting patterns account both for demographic and behavioral variation in the underlying dataset [1].

After applying NMF, we choose the representative demographic group for each latent factor based on the factorial weights from the NMF computation. For a given factor, the demographic group with the highest NMF weight is chosen by APG as the representative demographic group of the corresponding persona.

Finally, these representative demographic groups are enriched with additional information (picture, name, favorite games, job, relationship status, education status, and audience size) to create more complete persona profiles. The persona name and picture are chosen from a proprietary database that contains thousands of portrait pictures tagged for age, gender, and country [2]. Once a name or picture has been used for a persona, it will not be selected when generating new personas. This avoids the conflict of a persona having the same name or picture within a generated persona set.

**Generated Player Personas**

We generate 10 player personas using APG, corresponding to the generally low number of recommended personas [5]. An example of the automatically created player personas is shown in Figure 5, while Table 3 summarizes the set of 10 personas generated from the game preference data. End users of the APG can use the interface to select the generated personas (see Figure 4).

![Figure 4: The APG system (https://persona.qcri.org) provides an opportunity to browse and search the generated personas](https://persona.qcri.org)

![Figure 5: “Joonas” – an example of an automatically generated persona. The following information is retrieved using Facebook Marketing API: (1) job, (2) relationship status, (3) education status, and (4) audience size, and are matched with the demographic information in the survey data. The values shown for each persona are the most typical values of the Facebook segment corresponding to that persona, where the number of Facebook users is seen as an approximation of market size.](https://persona.qcri.org)
Table 3: Generated personas. The dataset had only Finnish respondents, so all the personas are Finnish as well – however, APG can capture geographically diverse audiences, as demonstrated in [18]. Platform is assigned based on the most dominant release platform of the personas’ favorite games.

Evaluation
To evaluate the accuracy of APG, we compare the generated personas with the baseline survey data along gender and age. Figure 6 shows that the generated personas correspond well with the gender distribution in the raw survey data. This is an important property, as it shows APG can capture the gender diversity of the players and thus help combat stereotypes such as gamers being male [9]. In terms of age correspondence, we use Kendall’s tau to compare the ranks of the seven applied age groups in the raw data and in the generated personas. The obtained value ($\tau=0.82$, $p=0.027$) shows high correlation, indicating that the personas follow the age distribution of the raw data well. This can also be seen from Figure 8 that shows the age group correspondence between the persona set and the baseline data.
system (see Figure 7). This information is important because it underlines the fact that different demographic player groups can “load” into the same persona based on the similarity of their game preferences. For example, Figure 7 shows that although “Male 25-34” is the dominant demographic group (weight=7.75), also “Male 18-24” (weight=0.04) and “Male 35-44” (weight=0.02) exhibit some game preference pattern typical for this persona. By examining the NMF weights that reflect these loadings, one can get a more granular understanding of different demographic groups’ game preferences.

Discussion

Contribution and Practical Implications
This research is a step forward in creating data-driven player personas. This is because previous attempts have produced less empathetic persona profiles, with missing demographic and other personified information that is seen important for user-centered design [10]. Specifically important is the fact that we are giving the player segment a name and a face (see Figure 5), rather than displaying only raw numerical information or abstract behavioral types [10, 13].

Showing player segments as “real people” is particularly important for marketing of video games, because these data-driven personas provide information on players’ demographics and preferences that gaming companies can utilize for ad targeting. For this, the AS feature (see Figure 9) is important because it helps to estimate the market size of the player segment that the persona represents. To illustrate this point, Figure 10 shows the audience sizes in the generated personas (M=130,400, SD=61,244 people).

Consequently, we outline the following use cases for APG-generated player personas:

- **Targeting player types:** As gaming ads are multi-billion-dollar industry, targeting players interested in specific games can decrease the cost of new player acquisition.
- **Persona-based recommendations:** Another potential application is recommending video games to players based on the persona they are most similar with, as contemplated in [19].
- **Reducing stereotypical thinking:** Using APG, game developers can create data-driven personas of players. This can help solve perceptual issues in the gaming community, such as gender stereotyping [9], as gender diversity in player types can be demonstrated using real data.

Limitations and Future Work

The presented “prototype player personas” still lack information of “fully rounded” personas [16], such as interests, pain points, and goals [5]. To provide useful representations for game developers and designers, future research should enrich the prototype personas with such information. In particular, we plan to use the results from a factor analysis conducted in a previous study [19] to enrich the prototype personas with game dynamics preferences (i.e., assault, manage, journey, coordinate, and care). This follows the idea of hybrid personas that combine the benefits of quantitative audience representation and qualitative in-depth user insights to generate holistic user representations [18].
References


