Inferring Interests from User Movements: The LISTEN-approach

Andreas Lorenz, Andreas Zimmermann
Fraunhofer Institute for Applied Information Technology
Schloss Birlinghoven
53754 Sankt Augustin, Germany
{Andreas.Lorenz, Andreas.Zimmermann}@fit.fraunhofer.de

Abstract

This contribution describes our experience in adapting audio presentations in a museum environment without falling back on explicit user feedback. It introduces the LISTEN project conducted by the Fraunhofer Institut in St. Augustin and partners, discusses results of the evaluation process and illustrates our approach of exploiting the user’s movement as the only interface for interaction. For this purpose, our contribution clarifies step by step how basic information like the user’s spatial position and head orientation, in combination with meta-information available for visual objects and sound snippets, are evaluated to conclude the user’s interest in the objects of the environment.

The LISTEN project deals with the audio augmentation of real and virtual environments. The system users move in physical space wearing wireless headphones and listening to a contextualized audio presentation. A realistic immersion of the user into the audio-augmented environment demands for the combination of both headphones, which are able to render 3-dimensional sound, and fine granular tracking information for the determination of the user’s exact position and orientation of the head. A first LISTEN prototype is applied to an art exhibition being launched at the Kunstmuseum in Bonn in July 2003. Visitors of this exhibition experience personalized audio information about art objects through their headphones.

Since the objectives of the LISTEN project emphatically disclaim the request of explicit user feedback, the utilization of implicit feedback is an important issue for the realization of a personalized immersive environment. The personalization process is based only on information about the spatial position of the visitor, the orientation of the head and some meta-information about sound snippets and exhibition objects. This meta-description consists of key-value pairs that hold information about colors, genre, epoch, etc.

The physical environment the user moves through interacting with the system is represented in the space model of the LISTEN system. This model contains the geometric information about the environment and the spatial position of visual objects. Based on this knowledge, an augmentation layer defines areas (i.e. zones, segments, triggers) within the space model for virtually structuring the spatial environment. Active information objects and sound sources are attached to these areas. The augmentation layer filters the position and motion of the user by dividing the dimensions the user moves through (position and orientation) into meaningful constraints and deriving continuous parameters. The interpretation of the user’s physical movements leads to valuable information about the user’s location (i.e. the entered zone) and focus (i.e. the object catching the user’s attention). Thus, the system is able to determine
connections between the user’s physical position and any visual or virtual object that is attached to this location.

The interpretation of the user’s focus in association with time information allows a deduction of the users preferences or interests in the observed objects. This interest model is based on the time the user spends with a certain object and enables a ranking of sound pieces with regard to their relevance for the user. The user’s focus always points to a visual object, which is in turn associated with a meta-description based on key-value pairs (e.g. color = “blue”, theme = “landscape”). Now, the user’s interest model is continuously refined by taking advantage of the time information: the more time the user spends looking at one visual object, the more the meta-description contributes to the user’s interest model. Starting with an empty representation, this accumulation procedure generically builds a 2-depth tree. The nodes on the first level of this tree are named after the keys of all meta-descriptions that have been processed so far. The children of these nodes represent the accordant key-values and have a numeric value assigned that expresses the strength of the user’s interest in this topic. This numeric value is increased every time a meta-description with the appropriate key-value pair appears as input. To fine-tune the user’s interest model, the LISTEN system processes the meta-description of the user’s object of attention in a constant three-second interval. In succeeding steps this interest model supports the recommendation, filtering and collaboration process. Since this model is implemented in a domain independent way, it may be employed in several domains containing objects described by meta-information.

The LISTEN scenario offers three possible strategies of personalized behavior. First, a sound can be played, that provides different levels of information depth about specific visual objects, or a more general background sound. Second, a specific visual object may be recommended if it meets the users interests. In this case one of two implemented attractor sounds is played: either one sound that seems to be emerging directly from the recommended object or a curator talk moving towards the recommended object. Third, two users can be stimulated for collaboration (i.e. to exchange knowledge, to discuss about art, etc.) by having attractor sounds emerging from the appropriate locations.

The basic concept of the personalization procedure determining the mode of the system, are motion styles representing stereotypical user behavior in moving through an exhibition. The concept of stereotypes is well known in the field of user modeling to categorize users by their attributes or behavior. In our domain the user may be sauntering around, goal-driven or standing still, in the last case being either focused on a certain object or unfocused. These motion styles emerge from the analysis of the location and focus of the user and their evolution over time.

<table>
<thead>
<tr>
<th>Motion Style</th>
<th>Presentation Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sauntering around</td>
<td>Background music is played.</td>
</tr>
<tr>
<td>Goal-driven</td>
<td>Object-dependant audio information in less depth.</td>
</tr>
<tr>
<td>Standing, focused</td>
<td>Object-specific audio information.</td>
</tr>
<tr>
<td>Standing, unfocused</td>
<td>Location-specific background audio presentation.</td>
</tr>
</tbody>
</table>

The definition of these stereotypes results from the first review process of the LISTEN system; they are meaningful, easy to detect and to revise. The interpretation of the user’s motion style in combination with the location determinates the presentation style and facilitates the pre-filtering process of relevant sound pieces. If the user stands still focusing the object, object-dependent information is presented. If the visitor moves slowly not being focused on one specific object, a zone-dependent, more general presentation starts. Finally, the selection of one specific sound piece to be played depends on the user’s history of already
known sounds and on his/her personal interest model. The following table summarizes all motion-styles together with the presentation styles they activate:

If the presentation style is determined by the user’s motion style, a Case-Based Reasoning system (CBR-system) supports the decision on which strategy is to be executed. In a nutshell, CBR is the process of solving new problems based on solutions of similar problems of the past. This powerful methodology copies the behavior of human problem solving in everyday environments. In the LISTEN system the problem description of a case is a snapshot of current user context (location, focus, interest model, etc.) and the solution is the name of the adaptive behavior or strategy that has been invoked in this situation. In addition, every case contains a timestamp, so that the cases in the case base are causally linked. Altogether, the case base contains one case for each visual object and cases for context snapshots of all users. A snapshot of the user’s current context is used as a query for the CBR-system to retrieve similar cases (i.e. to remember similar contexts) of the past. The retrieved result is either the context description of a visual object, of another user or of the current user. In combination with the motion style this retrieval result is interpreted appropriately in terms of recommendation, collaboration or filtering.

During the evaluation process we have presented the implemented system to a variety of experts made up of museum curators, artists, and composers, who have appreciated the success of the synesthetic experience: the visitors enjoyed the combination of audio-visual perception and felt as the interaction with the real visual objects was augmented. Namely, curators appreciated the possibility to deliver content concerning the artworks in an innovative, enriched and less descriptive way.

Critical points in the domain model were noticed: the zones of interaction surrounding each artwork were sometimes too small, thus forcing the visitor to approach the artwork very closely. Besides, some visitors could not realize whether the changes in the audio virtual environment were due to their movements in the space or were part of the audio sequence. In particular in the case of overlapping zones, the boundaries of the object zones were hardly localizable by the user. In order to overcome these problems, we are working on making the zones more flexible, creating some “breathing zones” in which a sound is more attached to the user’s behavior in observing visual objects. In this sense auditory icons providing some landmarks in the virtual environment navigation are meant to be inserted into the audio presentation in order to make the user aware of the interaction with the environment.

At the beginning of July, with the opening of the LISTEN art exhibition at the Kunstmuseum Bonn, there will be the possibility to perform evaluation tasks on the personalization engine of the LISTEN system. Due to the fact that no explicit user feedback is provided, we expect the interpretation of the system decision to be difficult. The strong explanatory capability of the CBR-system and suitable data mining tools will support the analysis of the system behavior.