Audio Stickies: Visually-guided Spatial Audio Annotations on a Mobile Augmented Reality Platform

Tobias Langlotz[#], Holger Regenbrecht[*], Stefanie Zollmann[#], Dieter Schmalstieg[#]

[#]Graz University of Technology
Inffeldgasse 16, 8010 Graz, Austria
{langlotz, zollmann, schmalstieg}@icg.tugraz.at

[*]University of Otago
P.O. Box 56 9054 Dunedin / New Zealand
regenbrecht@ims.tuwien.ac.at

ABSTRACT
This paper describes spatially aligned user-generated audio annotations and the integration with visual augmentations into a single mobile AR system. Details of our prototype system are presented, along with an explorative usability study and technical evaluation of the design. Mobile Augmented Reality applications allow for visual augmentations as well as tagging and annotation of the surrounding environment. Texts and graphics are currently the media of choice for these applications with GPS coordinates used to determine spatial location. Our research demonstrates that the use of visually guided audio annotations that are positioned and orientated in augmented outdoor space successfully provides for additional, novel, and enhanced mobile user experience.

Author Keywords
Augmented Reality; Spatial Audio; Mobile phone;

ACM Classification Keywords
H5.1. Multimedia Information Systems: Artificial, augmented, and virtual realities.

INTRODUCTION
Smartphones are ubiquitous in today's lifestyle, with increasing numbers of people using them to communicate with each other, or for browsing the Internet, or listening to music while walking in the city. Mobile technologies are also used to access passive location-based services (e.g., tourist information systems) and actively to leave tags or geocached items for instance. Augmented Reality (AR) can further enrich user experience by visually enhancing location-based services. Visual AR combines digital information with the real world by overlaying information in real-time, usually on captured live video streams.

A key characteristic of many AR applications is precise and fast tracking, which enables accurate augmentation of the displayed information. This capability is usually achieved by analysing the camera image for known features that can be tracked or by utilizing hardware sensors such as GPS, compass, accelerometers or gyroscopes, or by combining them. While previously, only desktop and laptop computers were able to deliver the required performance for AR applications, smartphones are now just as capable of running such applications with sufficient speed. This advance has led to the development of so-called AR browsers, such as Layar[1], Wikitude[2], and Argon[3]. Though they can be considered as Web browsers, such tools can also display digital content in real world contexts; a possibility that was foreseen by Spohrer (Spohrer, 1999).

The same time as people started developing visual AR applications, researchers have also built prototypes that use audio information or soundscapes to augment the immediate environment (Bederson, 1995; Rozier et al., 2000). Like visual AR applications, these prototypes began life as desktop-based technologies before their use on mobile devices such as smartphones.

But while precise tracking in 6 *Degrees of Freedom* (DoF) is omnipresent in the domain of *visual* Augmented Reality, most of the *audio* AR applications only use 3DoF or 2DoF tracking (such as GPS) to link the audio content to specific locations. The augmented audio information is only roughly placed, oftentimes rather resembling a positional hint than being precisely linked to (smaller) objects or particular locations. Despite the lower accuracy, several Audio AR prototypes such as (McGookin et al., 2011; Rozier et al., 2000; Woo et al., 2006) showed that for many audio AR applications a less accurate anchoring and tracking approach is sufficient to achieve a convincing augmentation.

However, in a similar way a higher precision is very desirable and needed, for instance in the case of many audio sources (Magnusson et al., 2010): being able to exactly position audio messages in a visually-guided, spatially correct way would extend the widely used principle of using textual or graphical sticky notes with the new concept of audio notes allowing novel and unexplored use-cases.

In addition, most existing applications rely exclusively on either visual or audio information for the augmentation of the environment, but do not consider their combined potential. Consequently, this paper, presents the use of *Audio Stickies* as a novel way of implementing spatial audio augmentation on mobile devices. Audio Stickies

OZCHI'13, November 25–29, 2013, Adelaide, SA, Australia.

[1] http://www.layar.com/

[2] http://www.wikitude.com/

[3] http://argon.gatech.edu/

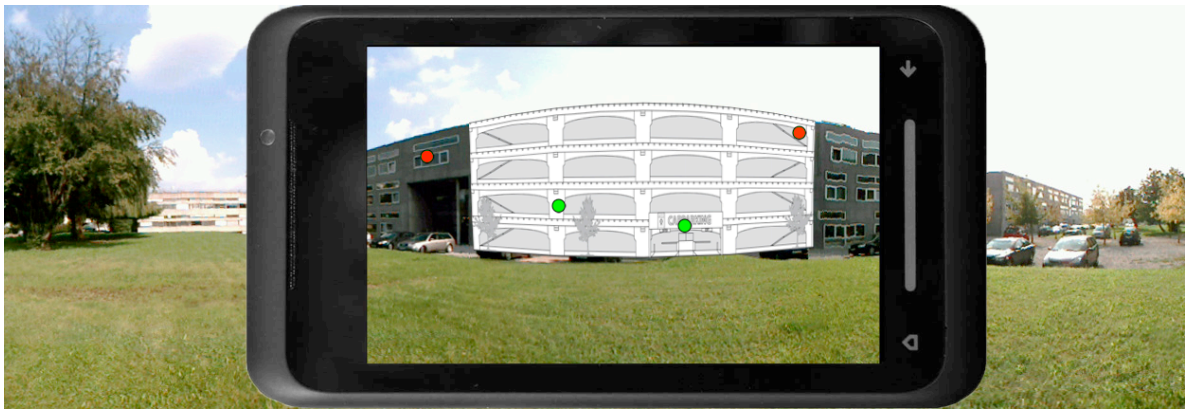

Figure 1: Concept image of the Audio Stickies browser: an architectural design alternative is spatially laid over an existing building. Colored dots indicate the position of user-created Audio Sticky comments that can be heard by pointing the center of the phone's screen towards them. The green dots indicate that the current Audio Sticky comment is played, while the red dots indicate that they are currently out of focus.

are user-generated spatial audio annotations that are precisely linked to the environment. Once created, the Audio Stickies can be acoustically perceived and controlled by looking towards them. By combining Audio Stickies with visual cues augmented in the users view, we guide the users and support control of the currently played Audio Stickies (see Figure 1).

We suggest a wide range of possible applications for this type of approach, in particular when used in combination with or substituting textual and graphical tagging. Audio Stickies can be placed effortlessly in indoor or outdoor spaces - they do not require typing, drawing or commands and therefore can contribute towards more natural, pervasive user interfaces.

The work presented here is based on an urban redesign scenario in its planning stage. Interested citizens use Audio Stickies to record their comments on various buildings in their pre-built planning stage. The Audio Stickies are placed directly onto the to-be-built, virtual architecture. This would allow for a more direct, natural and immediate way of participation in urban planning processes.

In sum, the project involved: (1) the development of a novel concept and implementation of Audio Stickies allowing users to create and share precisely placed spatial audio annotations (Audio Stickies), (2) the combination with augmented visual cues for guidance and control, and (3) a user study to evaluate the usability and perceived usefulness of Audio Stickies as a novel interface type. To our best knowledge, none of these aspects have been presented before.

The research that follows contributes to the fields of pervasive computing and can inform the design of future AR browsers and lends itself to end-user participation. It can also be used to create feedback systems based on spatial audio comments. A longer-term goal is to explore how we can interact with our surrounding environment in ways that maximizes the amount and quality of user-generated content with low user effort.

RELATED WORK

Research investigating augmented outdoor environments has involved three main approaches to date: (1) those relying on visual overlay of the environment to display additional information, (2) those focusing on Augmented Audio to provide additional information, and (3) those that combine elements of both (1) and (2). In 1997 Feiner et al. undertook pioneering work in the field of outdoor Augmented Reality using visual overlays as part of their MARS project (Feiner et al, 1997). This system used a backpack laptop computer combined with an external GPS and compass system that enabled the overlay of textual and graphical information on a campus information system. Later research developed interfaces to create textual annotations that also relied on the use of laptop computers (Wither et al., 2009).

By 2008 more and more smartphones were equipped with built-in sensors for estimating the position and orientation of the device, making mobile technology an attractive platform for outdoor AR. Recently, commercial companies like Wikitude and Layar have begun work on commercial applications that are technically similar to the Feiner et al. solution, but which are able to be implemented on smaller and widely available devices.

However, the accuracy and precision of the sensors that are integrated in smartphones is insufficient for high quality augmentation. Langlotz et al. have suggested merging the position and orientation estimate from the built-in sensor with the estimate obtained from a visual tracker analyzing the camera image (Langlotz et al., 2011). This work is based on an integrated system utilizing smartphones to place and share textual information in an augmented environment. They achieved a higher precision in tracking robustness and accuracy, compared to systems that rely exclusively on built-in sensors only. Langlotz et al. later demonstrated also the applicability of this system for video information (Langlotz et al., 2012).

For some audio applications, a less accurate tracking approach is sufficient. Bederson et al's museum tour

prototype automatically detects the current position of users based on infrared emitters in the museum (Bederson, 1995). A Sony MD player provided audio information that corresponds to the object being observed. Similar projects have used active badges for tracking the position of individuals in office environments. Audio information about the people is played when a visitor enters the office (Mynatt, 1997).

Augmented audio in outdoor environments has also been explored, where pre-recorded audio is played depending on specific GPS positions (Rozier et al., 2000; Woo et al., 2006). Spatial narratives have been used as a medium for location-based mixed reality (Dow et al., 2005).

Similarly, McGookin et al. aimed with their PULSE system an approach for an auditory display to geo-tagged social messages (McGookin et al., 2011). They used a text-to speech engine to play messages for instance from Facebook or Twitter.

Magnusson et al. presented a system for non-visual orientation and navigation (Magnusson et al., 2010). Unlike many other existing Audio AR systems that require the user to physically move toward a position to experience the audio augmentation, they proposed to use the pointing metaphor in a way that by pointing the phone towards certain directions audio augmentations are played to indicate directions.

While previous Audio AR systems seemed to not suffer much from the poor sensor-based tracking, Magnusson et al. stated that the error-prone orientation estimate from compass and accelerometers was a major drawback in their research.

A major problem in audio augmentation is the issue of overlapping sound sources. Vazquez-Alvarez's user studies showed that two sound sources playing simultaneously could be perceived separately, albeit at the cost of an increased mental workload. This workload is intensified in dynamic environments, where the users or sound sources are moving (Vazquez-Alvarez, 2010).

Usually, systems use either audio augmentations or visual augmentations. Approaches combining both systems are rare. Behringer et al. have implemented a system for device diagnostics, which uses augmented instructions together with a speech interface and audio comments that give further instructions (Behringer et al., 1999). Haller et al. use combined markers with 3D sound sources in a predefined setup for a more intuitive perception of the 3D sound in an indoor application (Haller et al., 2006). However, no user experience studies have been reported as part of this work.

Sundareswaran et al. have improved object localization by playing sounds at the position of the object concerned (Sundareswaran et al., 2003). They use a wearable setup similar to the one presented by Feiner et al. (Feiner et al., 1997) and extend it with sound capabilities. After a learning phase, users were able to detect objects based on sound more reliably.

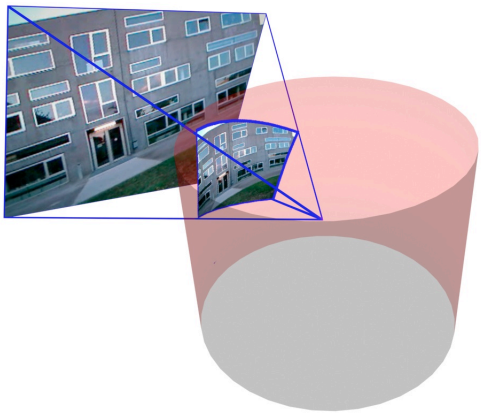

Figure 2: Projection of the camera image into the cylindrical-mapped panoramic image, which is used for precise vision-based tracking. A panoramic representation of the environment is created in the background, while the system is used.

Similarly, Rekimoto and Nagao implemented a prototype system, NaviCam, which detects color codes in the real environment as well as speech commands and synthesizes spoken messages (Rekimoto et al., 1995). The content to be played is pre-defined though.

To the best of our best knowledge, no systems have investigated precisely placed audio augmentations, which are created by the users. While (Rozier et al., 2000) involves user created content, it does not evaluate users' feedback or analyze the perceived usefulness of spatial audio comments. Like other outdoor systems, the approach is based on GPS position and compass orientation, rather than on precisely augmented sound. The reliance on GPS does not allow the users to place the audio comment in a sticky notes manner nor does it allow them to comment on smaller objects.

In this paper, we describe our contribution as exploring and evaluating user generated audio comments that are: (1) precisely and spatially linked to the environment, (2) visually guided, and (3) supported by visual augmentations in the context of urban planning.

VISUALLY-GUIDED SPATIAL AUDIO ANNOTATIONS

We present Audio Stickies as a novel way of implementing augmented spatial audio in an outdoor environment. Similar to written sticky notes, we wanted to place Audio Stickies precisely in our environment as a means of asynchronous communication. Users can leave Audio Stickies at certain, precise positions and other users can browse them by pointing their mobile phone towards visual hints representing the Audio Stickies. The visual hints are augmented in the user's view indicating that an audio annotation was placed there. To experience a seamless augmentation of the environment, precise and stable registration is crucial. This applies to visual augmentation as well as to augmented spatial sound and, therefore, also to our Audio Stickies.

Panorama-based Tracking

Reliable tracking in outdoor environments is still an open research issue, since GPS and compass systems, can be

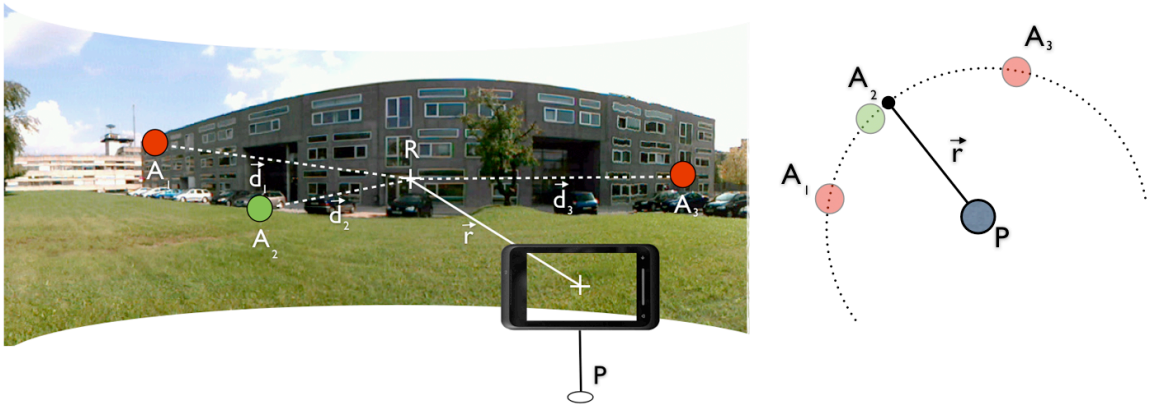

Figure 3: Illustration of the panorama-mapped Audio Stickies: (Left) User at position *P* browses Audio Stickies (A_1, A_2, A_3) in the environment. The current focus point *R* is determined by casting a ray *r* from screen center onto the panorama of the environment. The volume and the position in the stereo channel of each Audio Sticky are determined by analyzing the vectors (d_1, d_2, d_3) pointing from the focus point to each Audio Sticky. (Right) Top down view illustration.

inaccurate by several hundred meters and tens of degrees from the actual position and orientation. However, there are some research prototypes that allow more precise tracking on mobile devices and outdoors. Unfortunately, these prototypes require an existing model of the environment (Arth et al., 2011).

We chose to use a tracking system based on the work of Wagner et al. (Wagner et al., 2010). This system does not require pre-existing information about the environment, but uses a cylindrical-mapped panoramic representation of the environment that is created in a background thread (see Figure 2).

While being constructed and extended in the background, the panorama can be used for vision-based orientation tracking. It requires the user to (roughly) keep position while using the system, as only rotational movements are supported. Combined with the integrated GPS sensor, the system can be extended to support multiple positions. An absolute orientation estimate can be determined by further merging in information from the internal compass (Langlotz et al, 2011).

Spatial Sound with Audio Stickies

Using a panorama tracker, we built an application with an added capability that would record and play back Audio Stickies.

The user is "looking through" the mobile device and experiences an Augmented Reality view of the current environment. To create and place an audio annotation, the user needs to specify the point where she wants to place the audio annotation by touching the mobile phone's touch screen position. In the same way as textual annotations (Langlotz et al, 2011), Audio Stickies are stored in relation to a panorama coordinate system leading to a pixel-precise placement within the panoramic-space. Therefore, we had to transform the currently selected screen coordinate via the current tracking information into the corresponding coordinate in the panorama coordinate system. This can be achieved by casting a ray onto the cylinder that represents the cylindrical mapped panorama (see Figure 3).

Once the coordinate of the selected point is determined in the panorama coordinate system, a small widget is shown on the screen. The widget allows the user to record an audio comment. The recorded comment is then stored and referenced to the selected position. We limit the maximum length of each audio annotation to 10 seconds and interactively show the remaining time with a progress bar. Once created, the Audio Sticky can be shared and browsed by other people visiting the same spot, in the same way one can browse textual annotations (Langlotz et al, 2011).

To activate and perceive the Audio Stickies the user browses the AR view by moving the mobile phone. In each frame, we cast a ray *r* from the center of the screen - via the panorama cylinder – into the panorama to compute the focus-point *R* (the center of the currently visible camera image) of the user in panorama coordinates (see Figure 3). Once *R* is determined, we compute the direction vector d_n and the distance from all Audio Stickies A_n to the focus-point *R*. We play only those sounds for which the distance to the focus point is below a certain threshold (see Figure 3). Depending on the threshold, it is possible that several Audio Stickies can be played simultaneously.

We use the distance and direction to the focus point from the Audio Stickies to adjust the volume and position in the stereo channels. Consequently, Audio Stickies closer to the focus point – the screen center - play louder. Moreover, the position in the stereo channels corresponds to the position on the screen. Audio Stickies placed to the right of the focus point appear louder in the right stereo channel. Based on the recommendations in (Vazquez-Alvarez, 2010), we adjust the threshold so that only up to two sound sources are played at maximum volume in order to suppress for audio clutter.

We also make use of additional visual cues to guide the user's view to Audio Stickies in his/her current environment. We are augmenting visual dots at the position of Audio Stickies. Along with the visual guidance provided by the dots, they also support the control of the Audio Stickies playback. Audio Stickies that are currently playing are shown with a green dot, while inactive Audio Stickies - based on their distance - appear as red dots in the user's view. Once the user looks towards a red dot, the dot turns green and the Audio Sticky starts to play in a loop.

Implementation

The entire prototype application was implemented using the Studierstube ES framework presented by Schmalstieg and Wagner (Schmalstieg et al., 2008). Studierstube ES is an Augmented Reality framework optimized for mobile devices such as smartphones. We extended the framework to support sound recordings and the playback of spatial sound.

While implementing sound on a stand-alone, non-AR PC system is a rather trivial task, we accommodated a number of special requirements to run it on a smartphone. First, a sound engine is required that works with the limited capabilities of a smartphone. This requirement includes 3D support or at least controllable stereo sound. The sound engine should be able to play several sounds simultaneously, because multiple users may wish to place multiple Audio Stickies in the environment. Playback delay should not affect interactive real-time performance and the engine should have a relatively small memory footprint, allowing program and multiple sound data to be hold in memory at the same time. Finally, it should support certain sound file formats to achieve a trade-off balance between size and quality.

Several sound engines, audio storage formats/audio codecs and audio qualities/bitrates were considered and tested (native implementation on Windows Mobile, iAuxSoftSFX[4] and Hekkus Sound System[5]). Ultimately, we used the native API for recording sound and the Hekkus Sound System for playing sound, because the other options failed to meet our criteria (e.g. iAuxSoft has a big memory footprint). While the Hekkus Sound System is generally suitable, it does not fully support 3D Sound like iAuxSoft for example. We simulate 3D sound by using the sound panning between the left and the right channels and adjusting the volume settings. This technique is known as amplitude-panned sound sources (Ville et al., 2001). Even though the technique does not accurately simulate physical 3D sound, it is accurate enough for our purpose. We determined empirically that a sampling frequency of 27kHz was sufficient as the main purpose of the Audio Stickies was to record human voices. We stored the recorded audio files as WAV files, since playing several compressed (mp3, ogg vorbis) files

Figure 4: Participant conducting the user study in Dunedin: Participant actively browsing the environment for with augmented buildings and Audio Stickies from previous users using the mobile phone. Note the noisy environment next to a crowded street.

simultaneously caused a noticeable delay due to the limited computing performance of smartphones.

The final application performed with an average 25 frames per second on a Toshiba TG01 and on an HTC HD2.

USER STUDY

We tested the feasibility and usability of our prototype system and approach with a user case scenario in a controlled explorative field study, which is described in the following.

Scenario and Setting

There is a plethora of possible application scenarios. Almost all active, mobile location-based services, i.e. AR browsers and other applications where users leave text or graphics in-situ can be augmented with and benefit from Audio Stickies. Combined with social networking services, this facility could lead to versatile ways of

synchronous and asynchronous communication and collaboration.

To evaluate our system, we chose a case scenario involving public participation in urban planning**.** Instead of filing formal written proposals, our system allows the capture of immediate spoken feedback, hence improving public participation in the planning process. We suggest that this participation is best-achieved in-situ – where what is to-be-built can be viewed in its proposed context. Virtual architecture is visually overlaid over real urban scenes and citizens are asked to leave audio feedback on design alternatives.

In our scenario, we chose a location, which is to be redesigned as part of a bigger urban redevelopment, where existing buildings are to be demolished and replaced by new buildings. Public consultation regarding this type of project usually involves only textual descriptions illustrated with design sketches (if at all). Sometimes, if the building project is of wider interest or very large, crafted models made of wood and paper that people can comment on, or virtual flythroughs are provided in addition.

[4] http://www.iauxsoft.com/

[5] http://www.shlzero.com/

Figure 5: (Left) Participant of the user study while creating an Audio Sticky. (Right) Screenshot of the user interface showing existing Audio Stickies and an augmented building they can comment on. Tapping the screen at the designated position creates a new Audio Sticky. Created Audio Stickies are rendered as colored dot indicating the position and current state together with acronym of the author. The upper controls show can be used to control the program together with a panoramic representation of the environment that is created in real time and also displays Audio Stickies outside of the current view.

Our approach is to display new building designs in their actual context using Augmented Reality. We visually augment planned buildings onto the environment, overlaying existing buildings in order to give a more realistic idea of how new buildings will integrate in their environment.

Interested parties are given the immediate opportunity to provide feedback using our Audio Stickies approach. In this way, people using the system can comment on buildings, while they are still in the planning phase. The system also allows users to place their Audio Stickies precisely on the objects they want to comment on (e.g. elements of the façade). The whole system is implemented on off-the-shelf smartphones allowing for wide dissemination in the future, whether it is by experts or the general public audience.

For our usability study, we asked participants to browse augmented planned buildings that are visually augmented onto the real environment. Simultaneously, they were invited to use Audio Stickies to comment on particular parts of or on whole augmented planned buildings and to tell what they liked or even disliked (see Figure 4). Participants wore headsets (headphones with an integrated microphone) connected to a mobile phone, in the same way they might when listening to music on the go. In this way, they were able to listen to Audio Stickies created by previous users. Consequently, the number of collected Audio Stickies accumulated over time.

An initial pilot study with nine participants was conducted in a busy street next to the main campus of the University in Dunedin. The situation proved realistic and challenging with the noise of cars and pedestrians engaged in conversation. Noise affects the perceived quality of the audio annotations placed by participants. However, even in a relatively noisy location, the ambient noise was reported as being acceptable and we were able to use our prototype. Our pilot participants helped us to find flaws and positively commented on the general usability of the prototype.

Experimental Design

The user study was undertaken in two different environments: on the aforementioned busy street in Dunedin/NZ and in a contrasting quiet area on the university campus in Graz/Austria. This allowed us to estimate the influence of environmental noise and distractions on the usability of our system.

Fifteen participants were recruited for each site (30 total). None of those who took part were experts in Augmented Reality or in Augmented Audio. Twenty-two participants were male (73.3%) and eight female (26.7%); the age range was 21 - 50 years (M=28.44, SD=6.7).

Each participant was given a demonstration that explained how the prototype worked. They were then allowed to try the system and encouraged to ask any remaining question (see Figure 5).

Participants were then asked to browse four virtual 3D-models of the planned buildings for the site and allowed to listen to the Audio Stickies recorded by previous users. The models represented very different use cases of the buildings, including a car park, a food court, a teachers' college and a student accommodation house. An architectural designer created all the virtual building models that were used. We decided to create discussion-provoking designs to stimulate comments. The experimenter told the participants that the virtual buildings were possible candidates extending the current environment as part of an ongoing master plan.

Participants who wanted to record any of their comments using Audio Stickies were able to do so by touching the model (screen) at the appropriate position (also see Figure 5). The number of Audio Stickies increased with each subsequent participant. To initialize the system for the participants, the experimenter created two comments for each building design and used these as a starting point for discussions.

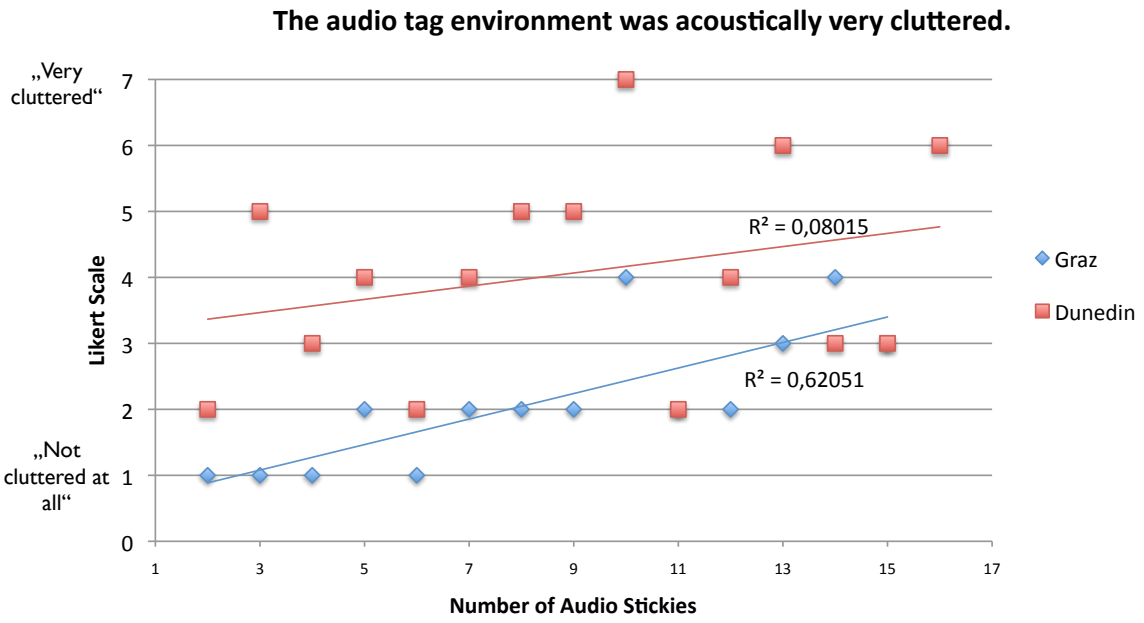

Figure 6: Feedback regarding the perceived audio clutter dependent on the number of Audio Objects with 1 = "Not cluttered at all", 7 = "Very cluttered".

While the participants were using the application, the experimenter noted any observations. The session ended when the participants had browsed all of the augmented buildings and did not want to place any more Audio Stickies. They were allowed to browse through the augmented building prototypes and audio comments back and forth for as long as they wanted.

After the participants finished commenting on the different parts and aspects of the building alternatives, they were asked to answer questions from a questionnaire using 7-point Likert-like scales ranging from 1 = "strongly disagree" to 7 = "strongly agree". The first part of the questionnaire contained demographic questions. The second part contained questions specific to the usability and usefulness of the prototype with items from a questionnaire developed by Lewis (Lewis, 1995) together with some questions specific to our scenario. The questions were followed by a short interview to try and elicit any potential problems or difficulties the participants experienced.

Results

All participants successfully finished the experiment and browsed the four proposed building designs together with the existing Audio Stickies. While they were not required to generate a specific number of Stickies, almost all participants created one Audio Sticky for each building.

Generally, the Audio Stickies were used to express user's opinions on certain aspects of the architectural design (e.g. "The planned facades of the building are mainly from concrete, which does not integrate well with the mainly green environment. Therefore, I wish that the architects rethink the use of more natural materials like wood or natural stone"). Other comments included the use of the buildings or the desire to see specific features (e.g. "I like the idea of adding a parking garage to the university campus, but I hope that the university also remembers to reserve some space in the building that can be used to drop bikes"). Others also used the Audio Stickies to comment on previous messages ("I agree to the other comments that the architects should use more natural materials and that the university should take care of existing and new green area").

The data gathered from the user study showed that audio annotations are seen as a useful source of information (M = 5.72, SD = 1.25, Figure 7(f)). While in Graz the average answer to the question "The audio tag environment was acoustically very cluttered" was 2.14 (SD = 1.03, Figure 7(h)) the participants in Dunedin scored it 4.07 (SD = 1.58). Scores for audio and visual clutter changed with the increasing number of sessions (and increasing number of Audio Stickies) for both test sites with different rate (also see Figure 6).

Participants in Graz answered the question "The ambient noise was very distracting" with 1.79 (SD = 0.80, Figure 7(m)) whereas participants in Dunedin answered this question different with 3.87 (SD= 1.64).

The participants answered that while using the system they could "easily identify the links between the audio tags and parts of the buildings" (M = 5.21, SD = 1.66, Figure 7(j)) but there was a difference in how easy it was to control (Graz M = 6.21, Dunedin M = 4.93, Figure 7(k)) and discriminate between them (Graz M = 5.57, Dunedin M = 4.60, Figure 7(l)).

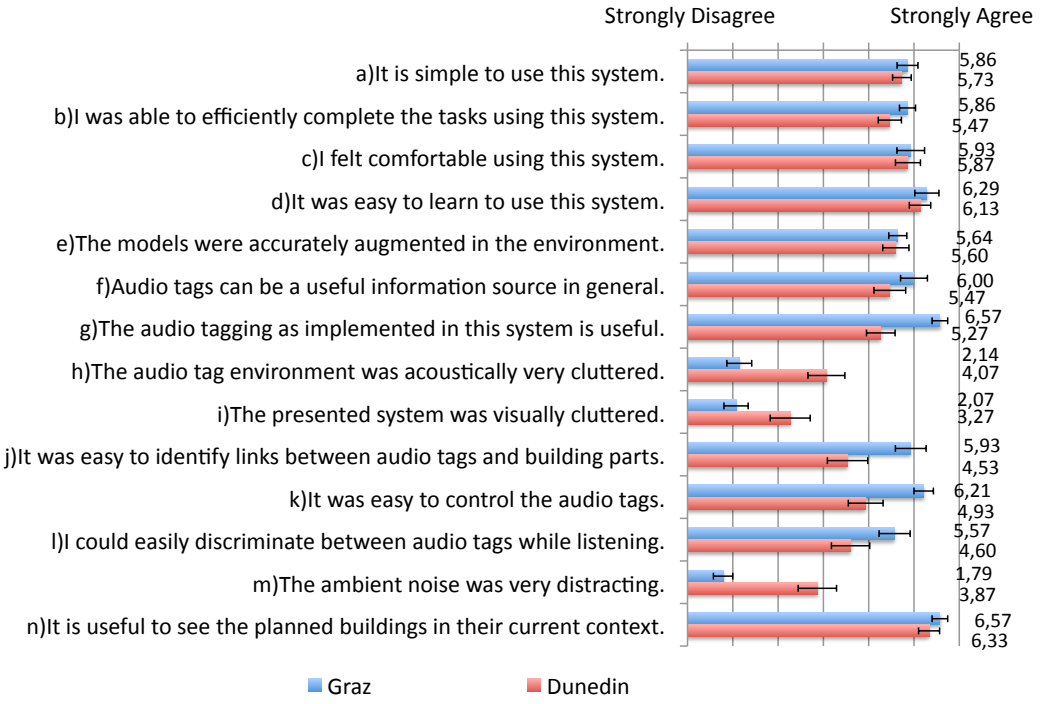

Figure 7: Questionnaire results of user study on 7-point Likert-like scales (1 = "Strongly disagree" to 7 = "Strongly agree").

Besides the differences between the two locations the participants agreed that the system was easy to learn (M = 6.21, SD = 0.94, Figure 7(d)) and easy to handle (M = 5.79, SD = 0.82, Figure 7(a)). This led to the consent that the participants solved the task efficiently (M = 5.66, SD = 0.86) (see also Figure 7(b)).

Audio tagging as implemented in this scenario was regarded as useful (M = 5.90, SD = 1.18, Figure 7(g)) and that seeing the planned building in their current context is very useful (M = 6.45, SD = 0.78, Figure 7(n)), supported by the stable and precise tracking expressed in "the models were accurately augmented in the environment" (M = 5.62, SD = 0.94, Figure 7(e)).

Several participants reported that they felt uncomfortable hearing their own voices as they did not like it – a feeling that many of us know from video tapes or audio recordings of ourselves and which is known to be a result of bone-conduction.

The increasing number of audio annotations did not seem to significantly affect the perceived visual or audio clutter, or the ability to localize the sounds or the perceived controllability or general performance of the system.

Discussion

The data gathered showed that audio annotations are perceived as useful source of information. Participants from both sites agreed that the system was easy to learn and easy to handle.

They also reported that the tracking was perceived as precise, stable and fast, allowing a seamless integration of the augmented buildings and Audio Stickies into the real world.

The participants from both locations reported that even if many Audio Stickies are present, it is still easy to control the annotations by looking towards them and to discriminate between different Audio Stickies. Regardless of the amount of Audio Stickies and ambient noise all participants were able to identify the link between the Audio Sticky and the object it relates to.

There was a difference between the two locations in terms of perceived audio distractions. In Graz, participants did not report on disturbing ambient noise. In Dunedin, however, all participants commented on how noisy the environment was.

While the results support our approach and the implemented prototype, it shows that audio clutter can still be a problem especially if many Audio Stickies are present. For both locations, the average results were acceptable but it was noticeable that usability of Audio tagging suffered a bit with an increasing amount of Audio Stickies and consequently the introduced audio clutter. The difference for both locations is likely caused by the different amount of environmental noise that adds to the audio clutter. Audio clutter seems to be a general problem in the domain of Augmented Audio and deserves more research.

Overall, participants positively reported on our approach of using precisely placed Audio Stickies as a general information source and as a natural way of interacting with the environment. We could demonstrate that Audio Stickies also work in rather noisy environments. The user interface was suitable even for novice users. The vision-

based tracking system worked seamlessly and did not cause any problems for the participants.

CONCLUSION AND FUTURE WORK

The paper presents a novel system for creating an Augmented Reality system based on user-generated and precisely linked augmented Audio Stickies. Accurate tracking allows the users to place audio information at a finer granularity, resulting in a higher accuracy and a higher density of Audio Annotations than with traditional (GPS) techniques. The contribution of precisely placed audio sources goes beyond quantitative measurements in a way that this precision is an enabler for many new applications scenarios of Audio AR requiring audio information precisely linked to small objects rather than roughly linked to areas due to sensor inaccuracies.

Consequently, the amount of audio sources in one's environment can be higher requiring new guiding and selection metaphors for selecting the audio information. In this work we showed how the system was combined with visual overlays to guide the users and highlight the position of audio comments and to display additional information.

We evaluated our approach with a user study, which allowed participants to express their opinion on proposed new building designs by using our approach of Audio Stickies. These Audio Stickies can be placed in the environment and are linked to real objects or augmented objects such as planned buildings or parts.

The user study demonstrated that audio annotations are seen as a valuable information source in general and users positively acknowledged the way they were implemented in our system. We were able to show that even inexperienced users were able to create, browse, and share audio annotations and that all users understood the link between the audio annotations and the objects they are referring to.

This work should encourage more researchers in the domain of pervasive computing and Augmented Reality to use audio – beside text and graphics information – as an additional and intuitive way of providing information to a user. The concept of Audio Stickies can be combined with existing approaches that display visual information without additional hardware.

By providing a prototypical implementation and proof-of-concept evaluation of the approach of Audio Stickies we have laid the foundations for future research that needs to target problems with the scalability of our approach and increases in audio clutter. The concept should be extended to a higher number of users which can operate the system in parallel or sequentially; The number of Audio Stickies per location and object will probably increase as well. While the current concept is appropriate for the current scenario and tests, chosen future settings might require the handling of a massive number of Audio Stickies. With an increasing quantity and complexity of audio annotations a "reply-to-message" mechanism would be needed or at least desirable. This would allow for more focused audio discussions amongst users. Filter mechanisms are needed to manage an increasing amount of Audio Sticky data, e.g. by time, thread of discussion, user, or spatial annotation position. The implementation of multiple-locations / multiple-users field studies will reveal the true potential and limitations of the concept.

Apart from improving public participation in urban planning processes, we hope our work contributes to the emerging field of AR browser developments by providing a different form of user content generation – enriching our surrounding environment by selectively sticking audio notes to it.

ACKNOWLEDGEMENTS

We would like to thank all users participating in the experiments. We especially thank Joshua Neary for providing the architectural models used during the user study and Abdulaziz Alshaer for his help on conducting the user study. Graham McGregor and Raphael Grasset commented on earlier versions of this paper. This work was partially supported by the EU funded project CultAR (FP7-ICT-2011-9 601139) and by the Christian Doppler Laboratory for Handheld Augmented Reality.

REFERENCES

Arth, C., Klopschitz, M., Reitmayr, G., and Schmalstieg, D. Real-time self-localization from panoramic images on mobile devices. 2011 10th IEEE International Symposium on Mixed and Augmented Reality, IEEE (2011), 37–46.

Bederson, B.B. Audio augmented reality. Conference companion on Human factors in computing systems - CHI '95, ACM Press (1995), 210–211.

Behringer, R., Chen, S., Sundareswaran, V., Wang, K., and Vassiliou, M. A novel interface for device diagnostics using speech recognition, augmented reality visualization, and 3D audio auralization. IEEE Comput. Soc, 1999.

Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J.D., and Gandy, M. Exploring spatial narratives and mixed reality experiences in Oakland Cemetery. Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology - ACE '05, ACM Press (2005), 51–60.

Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. Personal Technologies 1, 4 (1997), 208–217.

Haller, M., Dobler, D., and Stampfl, P. Augmenting the reality with 3D sound sources. ACM SIGGRAPH 2002 conference abstracts and applications on - SIGGRAPH '02, ACM Press (2002), 65.

Langlotz, T., Degendorfer, C., Mulloni, A., Schall, G., Reitmayr, G., and Schmalstieg, D. Robust detection and tracking of annotations for outdoor augmented reality browsing. Computers & Graphics 35, 4 (2011), 831–840.

Langlotz, T., Zingerle, M., Grasset, R., Kaufmann, H. and Reitmayr. G., AR Record&Replay: situated compositing of video content in mobile augmented reality. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (OzCHI '12), (2012), 318-326,

Lewis, J.R. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 7, 1 (1995), 57–78.

Magnusson, C., Molina, M., Rassmus-Gröhn, K., and Szymczak, D. Pointing for non-visual orientation and navigation. Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI '10, ACM Press (2010), 735.

McGookin, D. and Brewster, S. PULSE. Proceedings of Interacting with Sound Workshop on Exploring Context-Aware, Local and Social Audio Applications - IwS '11, ACM Press (2011), 12–15.

Mynatt, E.D., Back, M., Want, R., and Frederick, R. Audio aura. Proceedings of the 10th annual ACM symposium on User interface software and technology - UIST '97, ACM Press (1997), 211–212.

Rekimoto, J. and Nagao, K. The World through the Computer: Computer Augmented Interaction with Real World Environments. Proceedings of the 8th annual ACM symposium on User interface and software technology - UIST '95, ACM Press (1995), 29–36.

Rozier, J., Karahalios, K., and Donath, J. HearThere: An Augmented Reality System of Linked Audio, Proc of ICAD '00 (2000), 63–67.

Schmalstieg, D. and Wagner, D. Mobile Phones as a Platform for Augmented Reality. Proceedings of the IEEE VR 2008 Workshop on Software Engineering and Architectures for Realtime Interactive Systems, (2008), 43–44.

Spohrer, J.C. Information in places. IBM SYSTEMS JOURNAL 38, 4 (1999), 602–628.

Sundareswaran, V., Wang, K., Chen, S., et al. 3D Audio Augmented Reality: Implementation and Experiments. (2003), 296.

Vazquez-Alvarez, Y. Designing spatial audio interfaces for mobile devices. Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI '10, ACM Press (2010), 481.

Ville, P. and Karjalainen, M. Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning. Journal of the Audio Engineering Society 49, (2001), 739–752.

Wagner, D., Mulloni, A., Langlotz, T., and Schmalstieg, D. Real-time panoramic mapping and tracking on mobile phones. 2010 IEEE Virtual Reality Conference (VR), Ieee (2010), 211–218.

Wither, J., DiVerdi, S., and Höllerer, T. Annotation in outdoor augmented reality. Computers & Graphics 33, 6 (2009), 679–689.

Woo, D., Mariette, N., Salter, J., Rizos, C., and Helyer, N. Audio Nomad. Proceedings ION GNSS 2006, (2006).