Århus, June 2007

cARtAble

- an augmented reality audio play table

Kasper Aae
University of Aarhus, Multimedia
Wiener bygningen 118
Helsingforsgade 14
8200 Århus N
Denmark
kasper.aae@post.au.dk

Mads Lykke
University of Aarhus, Multimedia
Wiener bygningen 118
Helsingforsgade 14
8200 Århus N
Denmark
u060524@daimi.au.dk

Vitus Vestergaard
University of Aarhus, Multimedia
Wiener bygningen 118
Helsingforsgade 14
8200 Århus N
Denmark
vitus@daimi.au.dk

ABSTRACT

cARtAble is a table-based Tangible User Interface which attempts to sonically augment children's play with cars in a non-intrusive way. It tracks the cars and the users via fiducial tracking using a webcam and the reacTIVision framework. The car's position and orientation data is mapped to car sound states and these sounds are positioned spatially by comparing the user's position and rotation to the car's position. The sound is generated through the Max/MSP programming environment and is played back to the user through wireless headphones. Furthermore the cARtAble has the functionality of including static objects and background sounds – exemplified in the prototype by city ambience and an audible positioned construction work site.

Although running relatively stable, the current prototype does not satisfy the robustness required of a system meant for children's play. The main problems lie in the tracking of the fiducials and relate to the camera and the lighting setup.

Author Keywords

Computer vision, Augmented Reality Audio, Tangible User Interface, Fiducial tracking/tagging, Amoeba tags, ReacTivision, .

INTRODUCTION

Children's play has been the domain of many *Augmented Reality* projects. In the search of challenging the possibilities within the state-of-the-art technology it might seem alluring for scientists to focus more on the technology than on the context of its use. In this way there is a risk of smothering the natural play with the use of advanced equipment and other awkward technicalities that are not part of children's everyday play. Projects described by e.g. Billinghurst et. al. and Woods et. al [1,2] show potential in regards of their technical solution and its future promises, but the usability from a child's perspective might seem cumbersome, when the initial technological fascination has eased off. Holding a handheld 3D visor will affect your use of e.g. a book, since your arms get tired over time. Therefore the information provided by the 3D graphics in the above mentioned examples is not available to you unless you hold the visor in front of your eyes.

In our opinion it is important to keep the context of the use and the phenomenon you augment in mind, in order to avoid having the technology interfere with the user's natural behaviour, which in our case is children's play.

Furthermore, in our opinion many Augmented Reality projects have a strong visual focus, and pay less attention to exploring the possibilities that arise from focusing on other modalities, such as e.g. hearing. As described by Birger Langkjær [3, pp. 22-23] visuals have the tendency of placing the user in the periphery of the action, whereas audio places you in the middle of the action, because of the nature of our senses. In this way a child wearing a pair of wireless headphones could be placed sonically in the middle of "the action" at all times without affecting his or her natural interaction possibilities to the same extent as a handheld visor showing 3D graphics. The tangible and visual interaction will remain as always when playing, but the augmentation emerges from the coupling of the sound heard through the headphones with the action taking place in front of you, as it always has when playing. We see inspiring possibilities in this scenario and would like to explore the possibilities of augmenting physical objects from children's traditional play with audio.

Problem definition

"How can children's playing with traditional tangible toys be augmented with sound in a non-intrusive and intuitive manner, and still maintain its immediate qualities?

Delimitations

In order to downscale this project we have chosen to limit children's playing to their playing with toy cars on the planar surface of a table. The potential target group for the prototype we have created is considered to be kindergarteners, i.e. children between approximately 4-7 years of age.

RELATED WORK

Here we briefly summarize some of the work that has inspired the project on both a conceptual and technical level.

"Emerging Frameworks for Tangible User Interfaces", Ullmer and Ishii

This paper looks at *Tangible User Interfaces* (TUIs) as an alternative to the traditional screen-based dominance of *Human Computer Interaction* (HCI). It looks at how physical objects can play a role as both physical representations of digital information as well as controls for these. It also presents some key characteristics of TUIs and discusses how they can assist in making the physical and digital worlds be part of a more coherent whole.

"SenseTable" MIT

A TUI platform for tracking multiple objects on a flat surface. SenseTable tracks position and orientation of the objects using electromagnetic sensing which, according to the authors, frees one from some of the typical problems found in computer vision based tracking, such as occlusion, change in light conditions and higher latency.

"Audiopad", MIT

A surface-based TUI application based on the SenseTable platform for musical performance that uses LC tags for tracking. The system tracks position and orientation via one to two LC tags in each tangible object. The data from the tracked objects is used to control sound samples, and graphically information is projected onto the table from above, showing information about the objects.

"MultilightTracker" (MLT), AU

A computer vision based colour tracking system that can track the light from up to four differently coloured Light Emitting Diodes (LEDs) simultaneously on a planar, semi-transparent surface. Its key features are low cost, simplicity, a latency of approx. 100ms and easy availability, as it requires only a recent PC with a webcam to work. MLT supports back projection for showing graphical feedback on the tabletop.

"reacTIVision", Universitat Pompeu Fabra, Barcelona

An open-source framework designed for table-based TUIs. The framework includes a standalone, cross-platform application for computer-vision based tracking of up to 90 unique visual markers (fiducial symbols) in real time, as well as a network transport protocol for transmission of position and orientation information of the tracked objects. The reacTIVision distribution also includes a collection of client examples for a number of programming environments, which allows for straightforward prototyping of TUI-applications based on the framework.

"reacTable", Universitat Pompeu Fabra, Barcelona

A multi-user, synthesizer based musical instrument with a tabletop TUI and back projection of graphical information. reacTable is based on the reacTIVision framework, and allows for musical collaborations both locally, using the same table, as well as remotely, using multiple tables connected via a network.

"Techniques and Applications of Wearable Augmented Reality Audio", Helsinki University of Technology

This paper gives an overview of the concept of Wearable Augmented Reality Audio (WARA), and discusses some of the requirements and problems when designing and implementing a WARA-system. The paper focuses on a pseudo-acoustic representation of sound, using headphones and microphones, by recording the sounds around a user, combining these recorded sounds with virtual sounds and then routing both the recorded and virtual sounds to the headphones.

IDEA AND CONCEPT

A tangible user interface for children

cARtAble is an audio augmented table for playing with toy cars. On top of the table there is a road map and some toy cars. Through a pair of headphones with a microphone attached the children are able to hear the cars, the micro-landscape and their own recorded, and potentially manipulated, voices while they are playing (see Figure 1, below)

The system will keep track of the children's position and orientation as well as the position and orientation of the cars and other sound emitting objects on the table.

The type of sound the cars make will reflect their state, e.g. type of car, forward and reverse driving, how fast the car is driving, etc. The distance and direction of the cars and other audible objects on the table compared to the position and orientation of the individual child will also have an effect on the sounds the objects make. E.g. the father away an object is, the lower the volume of its emitted sound will be, and if the child moves an object to one side or if the child turns his or her head, it will affect the panning of the object's emitted sound.

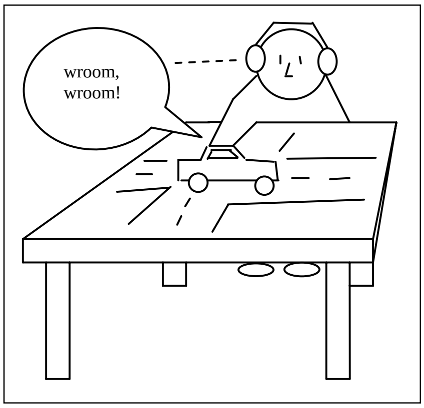

Figure 1 – a sketch of the basic cARtAble idea

Figure 2 – cARtAble was inspired by traditional toys like the Lego road tile-set and playmat roads.

In its essence the cARtAble is best understood as a TUI.

Using the interaction model and terminology by Ullmer and Ishii [4, pp. 3-4] the toy cars are both *physical representations* and *controls*. The physical state of the cars (position, speed, acceleration etc.) is thus reflected in the digital state of the system.

The only *digital representation* is the recorded and virtually generated audio, which is manipulated in real-time. The goal is for the system to produce audio that convincingly "belongs" to the toy cars, by having the user perceptually coupling the audio output to the physical state of the tangibles[1]. There will be no visual feedback projected onto the table, which looks like any other regular play table.

[1] Based on the thoughts on terminology by Ullmer and Ishii, we use the term "tangibles" to describe the toy cars as the physically manipulable artifacts of the interface

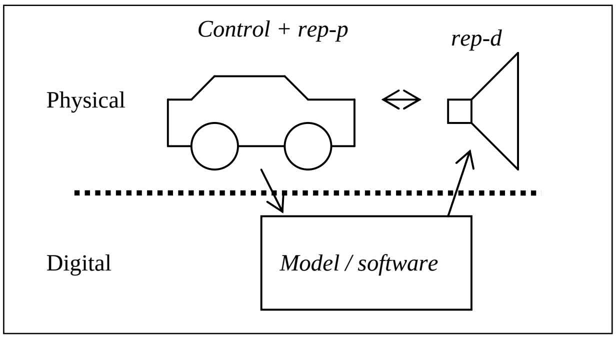

Figure 3 - The MCRpd[2] model of the cARtAble concept

The digital layer, as shown on Figure 3, should appear invisible to the user who should instead experience a direct link between the physical -and the digital representation (as suggested by the arrow between the toy car and the sound emitter).

Conceptually we are dealing with a *static binding* between the physical and the digital representation [4, p. 8]. We do not want the user (a child) to be confused, so a *dynamic binding* where the user can choose between several different sound schemes for the same physical toy car is out of the question. The cARtAble thus aims for the lowest level of user abstraction as possible; a fire truck should always sound like a fire truck.

The cARtAble as a system of objects

Consider the following ideal scenario:

The user walks around the table playing with cars on top of it. On the table the user can see a landscape with roads and buildings. He can hear each car relative to his own position and orientation and the same goes for the landscape where he is able to pinpoint stationary sound sources such as construction work, a car wash and so on. The toy cars respond sonically to being moved around, and the engine sound reflects whether the car is accelerating, going reverse, skidding sideways etc. When two cars collide, a crash is heard, and when a car stops at the gas tank the user can hear the car being refuelled.

The cARtable is a spatial and relational system [3, p. 9-10].

The system uses a *spatial approach* as it maps the position and orientation of both the user and the tangibles to the virtually generated audio.

But the system is also *relational*: It interprets the state of a car relative to the other cars or relative to locations in the landscape. The interpretation may result in an event, e.g. a collision of two cars.

[2] "Model-Control-Representation-physical/digital"

A creative mind could also come up with a *constructive approach* such as the addition of different exchangeable engines for a race car, different sounding trailers for a truck etc. But due to the goal of keeping the interface simple and understandable for small children, this kind of modularity is not a part of the cARtAble concept at the current time.

DESIGN AND IMPLEMENTATION

Although we do not claim to be experts on designing for children, we still feel there are certain criteria, the system must satisfy because of the target group. In the design we make the following assumptions:

- The physical setup should to be suitable for children. It should to be small and should not require e.g. a long reach or adult sized hands.

- The system must be robust and should be able to handle creative and unintentional use.

- The system should make sense at a very low level of abstraction, which means that the users should be able to make a logical coupling between what is happening on the table and what they are hearing.

-The hardware and other technical aspects of the underlying system should not be immediately visible to the user, in order to avoid interfering with the natural play of children. The table should look like any other play table.

Prototype Features

The cARtAble prototype is a somewhat downscaled version of the overall concept. The design and implementation of the prototype aims at clearly showing the core concepts of the cARtAble, rather than adding and perfecting a number of additional features.

The features of the prototype are:

- A rectangular play table with roads
- Tracking of one toy car (a truck)
- Relative sound positioning of the toy car compared to the user's position and rotation
- Sound that reflects forward and reverse driving as well as sideways sliding.
- Sound that reflects the speed of the car.
- One landscape with a non-interactive soundscape.

Features NOT implemented in the prototype:

- Head-tracking of user (instead the prototype features user movement simulated in the software that can be set manually).
- Support for multiple users.
- Recording of user's voices.
- Relational interaction between objects on the table.

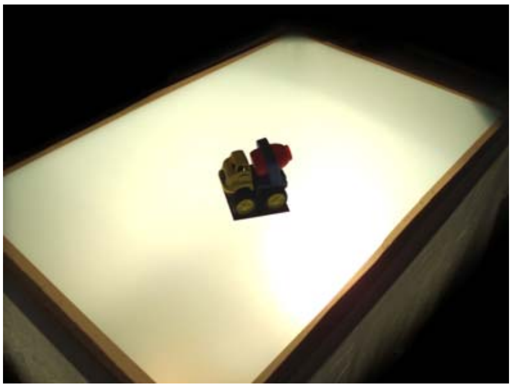

Figure 4 – the cARtAble prototype

System overview

Before we go into detail about the design choices and actual implementation details of each component of our system, we present and overview of all the components and how they are interconnected (see Figure 5).

1: car is moved around on the table by user.

2: a webcam placed under the table films the car.

3: tracking engine acquires video stream from webcam.

4: audio engine acquires position and orientation data of the car which is calculated by the tracking engine.

5: audio engine calculates audio data and sends it to headphones.

6: user hears the audio playing in the headphones.

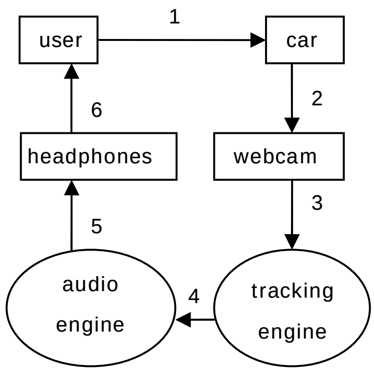

Figure 5 – overview of the components of the system and the flow of information between them

The Tracking Engine

The tracking engine is an important part of the system, since many of the design and implementation choices of the other components will be affected by which tracking method is used.

The key requirements that we have prioritized for the tracking part of cARtAble is a method that allows for rapid prototyping, as well as a system that allows for straightforward integration with the Max/MSP programming environment, as that is what the audio processing part of the system is going to be based on[3].

A webcam-based computer vision tracking system was the obvious choice for this. The hardware (the camera) was readily available to us, compared to e.g. an electromagnetic tracking system, like the one used in SenseTable [5], which probably would require some amount of custom construction. As a result of this, along with the requirements mentioned above, the reacTIVision framework has been chosen for the tracking. It supports accurate tracking of position, rotation and speed of the fiducial markers that are attached to the tangible objects (see Figure 6, below). It also comes with a basic input patch that allows Max/MSP to receive the output values from the tracking.

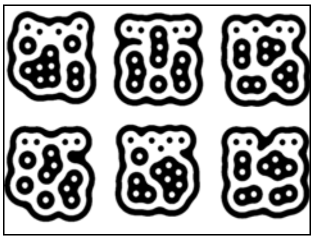

Figure 6 – The fiducial tags of reacTIVision, aka. "Amoeba Tags"

Before choosing fiducial tracking, we experimented with webcam-based colour tracking. This method was discarded, as the performance we experienced was less than favourable. Another reason for discarding it was that we would need at least two unique colours attached to every car in the system, in order to track both its position and orientation. A colour tracking project like MultiLightTracker [7] currently supports tracking of four different colours. Had we used a colour tracking system with the same features, it would be able to track only two different cars at the same time, which would make the system less scalable. The reacTIVision framework currently supports tracking of up to 90 unique fidicual markers, which means that the system will be sufficiently scalable, as each car needs just one unique marker to have its position and orientation tracked. Even though the cARtAble prototype only tracks one car, the scalability of being able to add additional cars and objects that can be tracked is still a priority for the overall concept.

In the cARtAble prototype, the reacTIVision application runs at a resolution of 640 x 480 at 15 frames per second using a Creative Webcam Live standard angle camera. The video stream acquired by the application from the webcam is thresholded into binary images to show only black and white, which are then searched for the unique black and white topologies of the fiducial tags. reacTIVision outputs ids of found tags in the image along with information about position, rotation and movement speed of the tags. This information is transmitted over the network using the TUIO protocol to the Max/MSP application which runs the audio engine[4].

The Table

The table setup is very dependent on the chosen tracking method as well as on the intended target group. Earlier we set up some design-guidelines, i.e. that the table should be suitable for children and that it should be able to hide the system hardware from plain view.

Ideally, the table would be round to facilitate and encourage the user moving around the table during play. However supposing that the child has an arms reach of 50 cm, the table surface would have to be less than 0.8 m^2 for the child to be able to reach the centre. This is not a whole lot of area to play with cars on roads!

By using a rectangular table the area can be larger, so this is preferable. The table should not be too high. The idea is for the child to stand up and walk around the table (as opposed to sitting on a chair) but even when standing, any table height above 1 m would probably exclude too many of the smaller children from using the table.

The webcam has been placed below the table, rather than above, in order to avoid the user accidentally occluding the camera's line of sight, and thereby interfering with the tracking. Placing the camera under the table also hides it from sight of the user, and it is shielded from much of the light in the environment around the table.

In addition to the shape and dimensions, the surface of the table also has to be considered. Since we have chosen to place the webcam under the table, the table surface needs to be transparent to some degree.

Having a table surface with full transparency makes it more difficult to control the light conditions, as light can shine into the camera from above, but also into the face of the user from below or reflect it back into the camera lens. It also increases potential tracking interference as the camera's line of sight goes beyond the table surface and thus giving the camera clear view of persons and objects that are above the table surface. A benefit from having a surface with full transparency is that the tags attached to the tangibles would not have to lie completely flat against the surface in order to be tracked properly. In this way, it would be possible to track a tag attached to the underside of a car, which would typically be a few centimetres above the table

[3] See the "Audio design and processing" chapter for more information

[4] See the" Audio Design and Processing" chapter

surface because of the car’s wheels. A fully transparent surface would also allow the car to be tracked even though the user lifts it from the surface. Whether this is desirable is up for discussion.

Having a table with partial transparency (and covered sides) allows much more control of the light conditions, which is vital when working with computer vision based tracking. Interference from light sources as well as other objects in the environment around the table will be minimized, and light sources placed under the table will not shine directly into the eyes of the users. With a table of this kind it would also be easier to hide the camera and other hardware placed under the table, which goes hand in hand with the design requirement of having the table look like any other play table. A downside of partial transparency is that the fiducial tags have to lie flat against the table in order to be tracked properly. This might cause problems, as the only elements of a car that usually touches the surface are the wheels.

With these pros and cons in mind, we have chosen to go with a partially transparent table for the prototype. Control of the light conditions, minimization of tracking interference and thus a potentially more stable and robust tracking was prioritized higher than having the freedom of placing fiducial tags higher than the table surface. This prioritization means that the design of the tangibles will be subject to certain restrictions (see “The Tangibles” chapter for more information).

The rectangular table used for the cARtAble prototype is 120 times 70 centimetres, and has height-adjustable legs that are set to a height of 90 centimetres. A webcam is placed under the table, pointing straight up at the surface and the sides of the table are covered with a white sheet. Two light sources (standard desk lamps with 40W bulbs) are placed under the table, on each side of the webcam, pointing inwards and slightly up. The white sheet helps reflect the light, in an attempt to create a more even and diffused illumination of the table surface to ensure a more steady tracking of the fiducials.

The Tangibles

The tangibles used in cARtAble are easily recognizable toy cars of different kinds (fire trucks, buses, sport cars, police cars etc.). In the brainstorming phase of the project we considered using symbolic tangibles such as coloured pucks. The advantage of that approach would be the implied *dynamic binding* between the physical representation (a puck) and the digital representation (sound) of the interface. The users would then have to bind a sound scheme to each tangible themselves. This would of course mean a more flexible (and perhaps fun) system, but realizing that we were designing for kindergarteners, we decided on keeping things simple and obvious by going with a *static binding*.

This means that the toy car used in the prototype visually is an easily recognizable truck, which has a static sound scheme bound to it that is equally recognizable. Since we are using a partially transparent table, the fiducial tag used to track the truck cannot be attached to the underside of the truck itself, but has to be attached to the underside of a platform onto which the truck is mounted. This means that the truck is not actually driving per se, as the wheels are not touching the table surface. This is quite an obstruction of how children would normally play with cars, and was spawned by the choice of using a partially transparent table. We will discuss this matter further in the “Evaluation and Discussion” chapter.

Figure 7 – images of the car used in the prototype, with a fiducial tag attached underneath

Mapping Between Tracking Engine and Audio Engine

The output from the reactTIVision tracking engine needs to be mapped into something that can be used to position audio in the virtual space and determine audio events. The audio needs to be positioned both in relation to the tangible’s position on the table, but also in relation to the user moving around the table. E.g. the sound from a car positioned to the right of the user needs to be panned to the right, and a sound from a car driving towards the user needs to have its volume gradually increased. If the user moves around the table in the outer perimeter, the volume and panning of the objects on the table should behave according to this as well (see Figure 8).

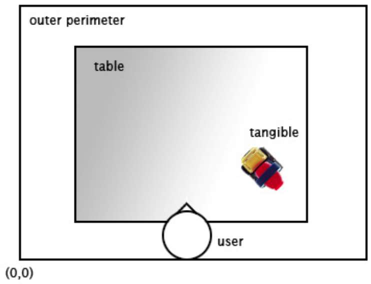

Figure 8 – The setup seen from above.

The panning and volume of the sounds is based on calculations of the distances between the tangibles placed on the table and the user's position, along with his/her rotation in relation to them.

The user's position and orientation is simulated in this prototype, based on values set in Max/MSP. These values can be changed dynamically, and is prepared for the actual tracking of the user in future revisions of the project.

ReacTIVision per default outputs the position (x,y), rotation and movement speed of the tracked fiducial tag.

The tangible's position is read as the (x,y)-values from reacTIVision. In order to avoid noisy values, the average of the latest values is calculated, and used as the car's current position.

This tangible's position value (index n) is compared to the previous position value (index n-1) to calculate the angle of the direction the tangible is moving in.

This angle is then compared to the rotation angle of the tangible, which is received directly from reacTIVision. In this way it is determined whether the tangible is moving in the same direction as it is pointing (= forward motion), or whether it is moving in the opposite direction (= backwards motion), and finally whether it is moving sideways compared to its rotation (= skidding).

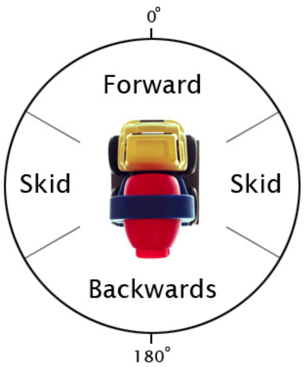

Figure 9 – The car's state is based on a comparison of tangible's rotation and its direction of movement

The three states (forward, backwards, and skidding) are defined via threshold values in relation to the comparison of the movement direction angle and the rotation angle. The areas that trigger the "Forward" and "Backwards" state are larger than the areas triggering the "Skid" state (see Figure 9)

Furthermore a stop state is needed when the car is not in motion. This state is calculated from the two (x,y)-pairs used for the direction angle. When these two pairs have the exact same values for a period of time, the car is considered to be in the stop state.

Audio Design and Processing

As mentioned above, in the chapter "The Tangibles", the sound scheme for the cARtAble has to be easily recognized and understood by the user (a child). Here we describe our design thoughts and the actual implementation of this in Max/MSP.

When playing with cARtAble, the sound is what makes up the actual augmentation of the child's play. The playing is described audibly through two layers of sound:

- a background layer, describing the setting (i.e. a city, a racing track, the countryside etc.)
- a foreground layer describing the user's playing with the car (driving forward, backwards, skidding etc.)

The background layer consists of cyclic *keynote sounds* and *sounds signals*[5]. These do not change according to the user's interaction.

The foreground layer describes the user's playing with the truck. The sounds playing are defined by the different states of the car (see description of mapping and possible states above).

We have chosen an urban setting with a construction site for the prototype of cARtAble. Because of this the environment consists of a city sound that loops all the time (keynote sound) and the sound of a jackhammer that plays once in a while (sound signal).

The keynote sound (city noise) is a sonic symbol of the urban setting. The actual sound is played back in stereo, and the level of the sound does not vary according to distances. Hence the level is always the same.

The sound signal (the jackhammer) symbolizes the construction site in the city. This sound is played back in mono, and can be placed anywhere in the virtual environment, using (x,y)-coordinates. The sound is not meant to be moved around by the user, but it is possible for us to place it wherever we want to, in order to create new environments.

The truck sounds played back are mono sounds like the sound signal, but its position in the virtual space changes as the user moves the car on the table. If the truck is moved to the left, the sound of the truck will be heard more from the left speaker. The actual sound playing depends on the trucks state:

- *Stopped*: an engine idle sound is played. This sound occurs whenever the tangible is not moved for a period of time. The functionality listens to the

[5] The terms keynote sounds and sound signals come from the Soundscape terminology, described by K. Wrightson [11, p. 10]. Keynote sounds are cyclic sounds that indicate a specific environment. Sound signals are event sounds that indicate a specific action.

x- and y-coordinate output from reacTIVision. If they do not change, the Stopped state is activated.

- *Forward*: a truck engine sound is played. It is started when the tangible is moved approx. forward in relation to its rotation. The speed of the tangible is reflected in the pitch of the engine sound. E.g. if the tangible is moved fast forward, the pitch of the engine sound is higher.

- *Skid*: a squealing tire sound is played, symbolising that the truck is moving sideways. It is played when the angle between of the tangible's rotation and its driving direction is approx. 90 degrees to either side.

- *Reverse*: a beeping truck reversing sound is played. It occurs when the tangible is moved approx. in the opposite direction of its rotation.

In order to determine the relative direction of the car (i.e. forward, reverse, skid/sideways) we simply compare the rotation angle of the car with the car's movement direction, as mentioned in the "Mapping" chapter. The rotation angle is delivered from reacTIVision through TUIO and the direction is calculated from position readings using simple polar coordinate system calculus.

But one thing is selecting the right sound to play – another is positioning the sound in the virtual space of the cARtAble.

To achieve a positioned sound we use the following techniques:

Panning: From the angle between the users head and the car we calculate a left-right panning. If the sound object is directly in front of or behind the user, we pan it to the middle so he hears it equally loud in both ears. The more to the left or right the sound object is, the more to the left or right we pan it.

Filtering: Because of the shape of the human ear and head, we hear sounds behind us differently than sounds in front of us. In short terms some of the sound frequencies are filtered, and in the cARtAble audio processing we try to simulate this. So the more the user turns his back to the different objects (e.g. the car or road work) the more we filter the sounds. We simulate distance the same way – by cutting some of the high frequencies and thereby approximating the way sound travels through air in the real world. In order to make it sound convincing we however exaggerate the filtering quite a bit.

Volume adjustment: When the distance between the user and a sound object is increased the volume is decreased. The amount of difference in loudness is exaggerated in order to provide a clearer distinction between close and far away objects.

The processed sound is routed from Max/MSP through the soundcard to a set of wireless headphones, in order to leave the user physically unrestrained.

EVALUATION AND DISCUSSION

User testing

We would have liked to involve actual users in the testing of the prototype, and we are confident that watching real children playing with the cARtAble would reveal unknown weaknesses and suggest alternative ways of playing with the cars.

The prototype however has never reached the degree of robustness where it would make sense to perform user tests with children.

If we were to test the table with a child, we would have to instruct the child not to go too fast, not to go all the way to the edges of the table, not to lift the car, not to lean on the table and so on. The number of restrictions on using the prototype is simply too high in our opinion to get any meaningful results from a test user.

The Camera

It is evident that the webcam is a weak link in our prototype. Its lack of wide angle makes it unable to cover the entire range of the table at the current height. It was a trade-off between the camera being as far away from the table surface as possible in order to be able to film a larger area while still being close enough to the table surface to track the tags properly. Having a frame-rate of 30 fps instead of just 15 would also greatly have improved the tracking performance of fast moving tags. Unfortunately the camera only supported 30 fps at the resolution of 320x240 pixels. This resolution would require us to lower the table in order to track the tags, and this would result in a smaller field of view. A wide-angle camera capable of running steadily at 30 fps in a resolution of 640x480, or even higher, would definitely improve the overall performance of the system.

The Light

The light setup of the prototype could also be worked on in order to improve tracking performance. Even though we managed to get the best out of the light available to us, we still found it difficult to get a completely even illumination of the table surface. This made it tricky to tweak the settings of the camera to work evenly well all over the table. In the corners of the table the tracking was somewhat unstable. This could be improved by using more light sources with diffusion filters in order to spread out the light.

The Table

The dimensions and height of the table works quite well as it is. It could be debated, whether the table area is suitable compared to the size of the car used in the prototype, but that is likely a matter of the car being too big, rather than the table being too small.

As for the surface of the table, it would definitely be an idea to test a more transparent type. It would allow us to overcome some of the design restrictions imposed on the car, but as we discussed earlier, it might also introduce some problems with regards to the tracking. We might lose some control over the light conditions and risk having more interfering objects in the image due to the camera having a longer line of sight.

The Tangibles

As mentioned above, there has been some substantial design restrictions imposed on the car used in the prototype. The necessity of mounting the car on a platform with a tag underneath severely hinders the toy car's natural behaviour, as it does not actually drive using the wheels. This goes against the requirement of having a non-intrusive system that does not alter the core mechanics of children's playing with cars, as the car in the prototype does not have a very car-like behaviour. As mentioned, this restriction could be overcome by using a more transparent table surface, as it would allow us to attach the tag directly to the car itself.

Also, the car may be too large for the table used in the prototype. This is mainly due to the quality of the camera which requires a fiducial tag of quite a large size in order to ensure stable tracking. Having a better camera would allow us to use a smaller tag, which in turn would allow us to use a smaller car.

The Mapping

The mapping between the four states of the car (idle/inactive, forward, reverse and sides) and the sounds generally works well technically and perceptually. It was a good choice to make the forward and reverse angles larger than the side-angles. Perhaps the angles that result in the skid-sound should be even smaller, so the car only skids when it moves "straight" sideways. The system needs to even out the incoming values more, so the car does not momentarily jump to an unexpected state. In the current system the car once in while e.g. gives a skid-sound when going forward because of tracking inaccuracy or because of the user simply shaking his hand a bit. Although this it not a problem in the logical design of the mapping between car movement and the sound output, the user will experience it as such. The user must never feel that the car is not responding correctly to his control – in that case the system is worthless. So getting better mean values for the movement of the car is very important.

The Audio

The sound positioning in the current system is not completely convincing. In order to achieve a higher degree of "externalization" – that is believably positioned sounds coming from external sources [10, p. 3] - there are a couple of techniques we could have used. Spectrum difference in the two ears using the "head related transfer function" would increase the sense of direction dramatically. In the current system we filter both ears equally and only discriminate between the ears through simple panning. The use of reverberation would increase the sense of distance. Actually this is a rather "cheap trick" that was not implemented in the prototype because we did not want to increase the CPU-load and was afraid that reverberation would make the sound too "muddy". In retrospect the system could probably have handled a lightweight reverb effect just fine, and if we did not over-use it, it could have added nicely to the user's feeling of distance.

The audio transitions between the states of the car could be improved a bit. In the current system there are some very simple transitions where e.g. the pitch falls smoothly when the car goes from a high speed to idle state. However the transition between e.g. forward driving and sideways skidding is not very smooth. Currently the car is either skidding or not skidding which results in a rather abrupt change in sound. A better solution would be to implement degrees of skidding either through direct mapping to different angles of movement or through interpolation.

FUTURE WORK

The cARtAble system is very scalable, and there are a lot of possibilities for expanding the concept.

Being a children's play table, it would seem obvious to make the table a multi-user system. Technically it would not require a lot of changes, since the system already has the functionality for tracking a user and generating the correct audio for that user's position and rotation. So with regard to the tracking, it would be a simple question of adding more user objects in the software. The biggest change would be the need to add a microphone for each user for communication. The microphone input would need to be filtered (i.e. turned into a "pseudo acoustic" sound environment) in order to blend with the virtual sound environment. A bonus would be the possibility to let the user's voice become a part of the game by e.g. tying it to a toy on the table and letting the child "be" that toy as in more traditional play.

An even more obvious expansion to the system is of course the addition of more cars and settings. It would be nice to have transparent roadmaps with fiducial tags that are automatically recognized by the system. The system would then play the background sounds for that setting (e.g. a parking basement or a seaside road) as well as adjusting the filtering and reverberation appropriately. The addition of additional tagged objects such as a church tower also seems a natural extension to the system. Later on other types of machines, or perhaps even animals could be added.

In order for the table to stay fun, the objects would probably need more states. A police car for example could have a siren. It would be a challenge to determine how to activate non-movement related states. The system already measures rotation acceleration, so one possibility would be to shake the car a bit to start the siren. The mapping however would probably not seem completely natural and intuitive.

For the game to really be rich on variety, interaction between objects would be a nice option. There is no problem in determining when two objects is close together, so the real challenge would be to make all the rules of interaction. Possibilities would include cars colliding, cars refueling at a gas tank and so on.

CONCLUSION

The overall aim of the project was to sonically augment children's playing with cars in a non-intrusive and intuitive manner, without removing the immediate qualities of that kind of playing. In addition to this, we also had some overall design requirements for the design of such a system: the physical setup should be suitable for children, the system should be robust and still work even when used creatively, it should work at a low level of abstraction and the hardware of the system should be hidden from plain sight.

Our system is based on the reacTIVision framework that tracks fiducial tags using a webcam. In our prototype, we have attached this tag to the bottom side of a platform onto which a toy car has been attached. The car can then be moved around on a partially transparent table, and be tracked by the webcam placed under the table. The tracked position and orientation data is mapped to different states for the car, which in turn are mapped to a specific type of sound, reflecting the car's behavior. The sound is positioned spatially according to the users relative position and orientation compared to the car. The audio engine is created in Max/MSP.

Some elements of the prototype could be improved. Regarding the tracking, it is evident that a camera of higher quality and additional and diffused light sources could be very beneficial to the overall robustness of the tracking but also make the system able to track smaller tags. This could also help the fact that the car is a bit on the large side compared to the size of the table. Experimenting with a fully transparent table surface, rather than the partially transparent surface used in the prototype, might also be able to fix the restriction of the car having to be placed on a platform, which hinders its natural movement.

In spite of these potential improvements, the system works as intended on a basic level by sonically augmenting the playing with cars. We have created a physical setup with a table that has dimensions that are suitable for children, the system works at a low level of abstraction by mapping car movement to logical car states that are mapped to logical car sounds and we have managed to hide the hardware from plain sight. The system is not as robust as we would have liked, and some details, like the car having to be attached to a platform, interferes somewhat with the how children would naturally play with toy cars – but with the above mentioned ideas for improvements, this could be remedied.

ACKNOWLEDGMENTS

Thanks to our teacher and supervisor Kaj Grønbæk, who also lent us the glass table. Thanks to Ole Caprani who lent us the webcam, even though we would have preferred a better one ☺ Thank you, Rasmus Lunding, for you patient and willing OS support and for letting us work in the basement with all the nice sound equipment. Finally, thank you Allan Nielsen, for you personality and for lending us the DV-cam and editing facilities used in the creation of the video demo.

REFERENCES

1. Billinghurst, M., Hirkazu, K., Poupyrev, I. *Projects in VR: The MagicBook - Moving Seamlessly between Reality and Virtuality.* 2001
2. Woods, E., Billinghurst M., Aldridge, G., Garrie, B. *Augmenting the Science Centre and Museum Experience.*
3. Langkjær, B. *Filmlyd & filmmusik – fra klassisk til moderne film.* ISBN 978-87-7289-336-5
4. Ullmer, B., Ishii, H. *Emerging Framework for Tangible User Interfaces.* In: Carroll, J.M. (ed.), Human-computer Interaction in the New Millenium, August 2001, pp. 579-601.
5. Patten J., Ishii H., Hines J., Pangaro G. *SenseTable: A Wireless Tracking Platform for Tangible User Interfaces.* Proceedings of CHI 2001, March 31 - April 5 2001, ACM Press, 2001 ACM
6. Patten, J., Recht, B., Ishii, H. *Audiopad: A Tag-based Interface for Musical Performance*
7. Nielsen, J., Grønbæk, K. *Vision Based Multi Object Tracking On Semi-transparent Surfaces*. Interactive Spaces, Dept. of Computer Science, University of Aarhus, Denmark.

8. Kaltenbrunner, M., Bencina, R. *reacTIVision: A Computer-Vision Framework for Table-Based Tangible Interaction*.

9. Martin Kaltenbrunner, M., Jord`a, S., Geiger, G., Alonso, M., *The reacTable: A Collaborative Musical Instrument* Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

10. Härmä et. al. *"Techniques and Applications of Wearable Augmented Reality Audio*. Audio Engineering Society, Convention paper, March 2003

11. Wrightson, K. *An Introduction To Acoustic Ecology*, The Journal of Acoustic Ecology Vol. 1, Number 1, 2000.