Spatial vs. Static sound - the differences it makes for 360° videos, highlighted by heatmaps

S. Rothe[1], M. Rieger[2]

[1] *Ludwig-Maximilians-Universität München, Germany, Email: sylvia.rothe@ifi.lmu.de*

[2] *VRTONUNG, Munich, Germany, Email: info@vrtonung.de*

Abstract

Cinematic virtual reality has been increasing in popularity over the last years. While watching 360° movies with head mounted displays, viewers can freely choose the direction of view, and thus the visible section of the movie. So, the viewer has a spatial visual experience, however, for the most 360° movies currently the sound is not spatial. For validating the influence of spatial sound, a user study was conducted, where the viewing behaviour for VR videos with spatial and non-spatial sound was compared. We used heatmaps of viewing directions and applied statistical analysis methods to spatiotemporal data. Using approaches of spatial statistics – analysis of space-time cubes and Getis Ord Gi* statistic – we found out that the behaviour for both conditions was often similar. However, in the spatial case, objects with sound attract the viewer's attention for a longer time.

1. Introduction

360° movies are attracting widespread interest and have many possible applications, e.g. telling stories about exciting locations in the world or ancient places of interest in history. Especially, museums and other educational institutions can take advantage of this as well as documentaries or travel experiences who bring the user to places, they can't go to.

In **cinematic virtual reality (CVR)** the viewer watches 360° movie using a **head mounted display (HMD)** or other VR devices. Thus, the viewer is inside the scene, and can freely choose the direction of view. Accordingly, the viewer determines the visible section of the movie – the **field of view (FoV).** Therefore, it is not always possible to show the viewer what is important for the story. Several conventional methods of filmmaking for guiding the viewer's line - such as close ups or zooms - are not practicable in CVR.

In our work we examine where the participants are looking at without any search tasks. We investigate if they follow several sound cues. In contrast to traditional movies, it is unclear if the source of a sound is in the FoV of the viewer, it depends on the viewing direction. In our experiments we compare the viewing behaviour for movies with spatial sound (Ambisonics, .tbe-file) and non-spatial (stereo .wav-file) sound. We are interested whether sound can attract the viewer's attention and if there is a difference watching a movie with spatial and non-spatial sound.

Non-spatial or static means that the sound always remains the same, even if the viewer moves the head and the field of view is changing. This sort of sound is also called head-locked stereo and here represents a downmix also named as static binaural stereo. So, the soundtrack for every user of this group always remained identical and is not depending on the head movement of the viewer.

Spatial means that the sound, just like the image, changes through head-movements of the viewer. The direction of the sound is identical with the source of the sound in the movie. Synonyms would be 3D audio, immersive sound, dynamic binaural stereo. The soundtrack was perceived as something different for each user. While the so-called 360° sound field was the same for each user, the sound for the current direction was rendered in real time depending on the head-movement of the viewer.

To explore this, we logged the head movements of the participants and visualized them in several ways. Similar to geodata, our logged data have two space coordinates on the surface of a sphere (from -90° to 90°, from -180° to 180°, fig. 1) and a time coordinate (timecode of the movie). The head direction is a vector (λ, φ) where λ represents the yaw and φ the pitch.

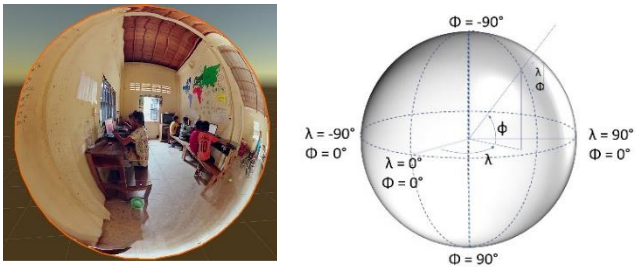

Fig. 1. The sphere of a 360° movie has the same coordinate system used for the earth.

For studying our collected data, we developed an analysing tool for movies which generates a heatmap for every timecode. Basically, there are several options for displaying the heatmap, e.g. via HMD or using flat displays. The perspective of the re-searcher, to explore the data on a sphere via HMD seemed unsuitable to us, since just a small part of the data is visible.

For presenting a full 360° image on a flat screen, the movie is represented by an **equirectangular projection** of the 2-dimensional sphere to a flat plane:

$$P: \begin{pmatrix} \lambda \\ \varphi \end{pmatrix} \rightarrow \begin{pmatrix} x \\ y \end{pmatrix}, \quad \lambda, x \in [-\pi, +\pi], \quad \varphi, y \in \left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$$

Our tool uses this approach and transforms the logged data in the same way. The projected data are visualized on the movie using **heatmaps**.

Using equirectangular projection for the movie and the data has the advantage of seeing the whole picture at once. However, the further an area is away from the horizontal 0°-line, the more the area is **distorted** (fig. 2).

Although we took the deformation into account when developing our heatmap tool, we assume that for our experiments there was no big influence of the distortion on the results, because head movements far from the horizontal line were very rare in our dataset. As can be seen in fig. 9, there were only few data outside the 20°-area. This was caused by the composition of the movie - all points of interest were near the horizontal line (0°) - and by the fact that we used head tracking and not eye tracking. Comfortable head movements range in a 30°-area only [1], whereas eye movements deviate further from the 0°-line. For this reason, we neglected the distortion in the inferential statistical analysis part of this paper.

Fig. 2. A manual generated heatmap for studying the distortion in the heatmap tool.

Heatmaps are suitable for finding clusters in the data, but they are not sufficient for inferential statistics to make statements about confidence. Therefore, **statistical methods** are needed for determining the significance of the clusters – for identifying hotspots.

The term **hotspot** is used in the literature in different ways. In our work, a hotspot is a cluster with high values identified by statistical methods using confidence levels. In our dataset, the value is the number of views in a space-time segment.

Our collected data are spatiotemporal data – data which have a space and a time component. This type of data is often used in geographical researches, so we applied some methods of spatial statistics which are used in geography: space-time cube analysis and hotspot analysis using Getis-Ord Gi* statistic.

2. Related Work

Much research in recent years has focused on presence in virtual reality environments – which can be adapted to CVR. Spatial sound leads to a higher level of presence [2,3] and increases the sense of place [4]. In our research, we focused on finding out where the viewer is looking at and if spatial sound changes the viewing behaviour.

In 1996 Pausch et al.[5] examined how the attention of the viewer can be drawn to a desired spot. The orientation of the head was logged, and conventional histograms were used for illustrating the data. The results show that most people hardly turned their head. The authors visualized this by histograms showing the rotation angle. Similarly, Sheikh et al. [6] connected several cues (motion, gestural and audio cues) to the main character of a scene. The head orientation was recorded and the percentage of people who had seen the target over time was evaluated. In their experiments, the cues with an audio component were proved to be more helpful than just visual cues, even if the sound was not fully spatialized. The results were displayed by diagrams showing the time for seeing the target. In histograms, tables and diagrams specific values were presented. With our approach, we want to illustrate the viewing behaviour in heatmaps and analyse the data using spatial statistical methods.

Spatial audio can be used for solving tasks in virtual and augmented reality. The use of spatial sound improves the results in search and navigation tasks [7,8]. Van der Burg et al. [9] showed that audio cues (pop) synchronized to a salient visual cue (pip) reduces the search time, even if the audio cue does not have any location information. Emil R. Høeg et al. [10] enhanced this experiment to virtual reality with sound cues from the same direction as the visual cue. They demonstrated that binaural cues lead to shorter search times, even though the visual cue was not always visible at the moment the audio cue was presented. In the experiments, the participants were given a search task in an abstract VR environment. In our study in comparison to [9,10], we move closer to a real cinematographic setting by using a realistic scene instead of abstract symbols and by not giving a concrete task to the participants but letting them choose freely what to do next.

To investigate the viewing direction for every timecode in the movie, we use space-time cubes (STC) and the Getis-Ord Gi*statistic. STCs were introduced by Hägerstrand in 1970 [11] (Hägerstraand 1970) and can be used for analysing geographical data [12–14]. The Getis-Ord Gi* statistic was established by Getis and Ord [15,16] for analysing spatial data. Songchitruksa et al. [17] used the Getis-Ord spatial statistic for identifying hotspots.

3. Analysing Methods

3.1. Heatmaps

In our experiments, the centre of the field of view (FoV) was recorded, and a heatmap was generated for every time interval in the movie. The heatmap represents how frequently viewers looked at a certain area segment within a specific time frame. For this we mapped the interior of a circle to each viewpoint and used a gradient function to compute the colour and the opacity of the overlapping areas.

Our heatmap tool is implemented in HTML5 using the HTML5 video element. This element has an additional HTML-tag: the track tag which was originally designed for text tracks defining subtitles. We used this track for generating the heatmap (fig. 3). A track tag requires a vtt-file (video text track). The recording script logged the head direction 60 times per second, which can be adjusted. The data were written in vtt-files.

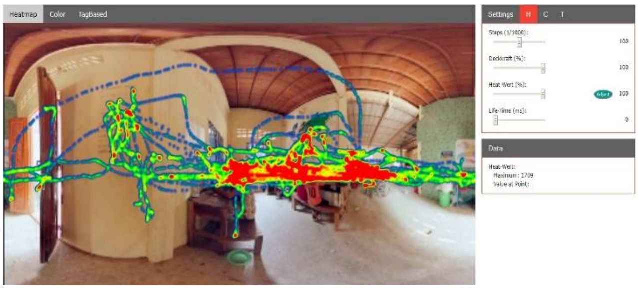

Fig. 3. View of the movie in the heatmap mode.

Investigating the heatmap, several parameters can be adjusted, e.g. the time interval, the cell radius, the scaling and the opacity (heatmap mode, fig.3).

Inspecting heatmaps is a first approach for finding conspicuities and clusters in the data. However, heatmaps do not provide any information about the significance of the results. For this it needs examinations using statistical methods.

3.2. Space-time Cubes

A method of investigating spatiotemporal data is to analyse space-time cubes (STC). In these cubes two coordinates represent the space and the third one represents the time. With this technique the data can be visualized and explored in a comfortable way (fig.4).

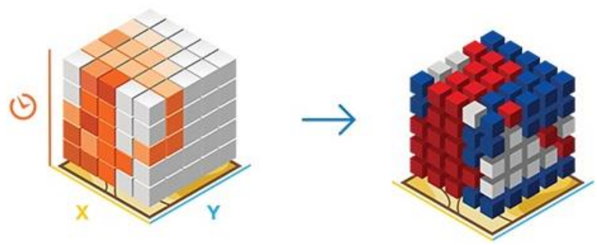

Fig. 4. A space-time cube has one time and two space coordinates, left: the density of a value is shown for every space-time section, right: the result of a hotspot analysis (significant hotspots are red) [18]

For calculating the STC and applying spatial statistical methods, the GIS software ArcGIS Pro was used. We converted the vtt-file into an Excel file which was imported into ArcGIS to employ the implemented STC method for analysing the data.

In a first step, we inspected space-time cubes showing the counts of incidents – in our case "how often" users looked at a certain area segment within a specific time frame, as shown in fig. 5.

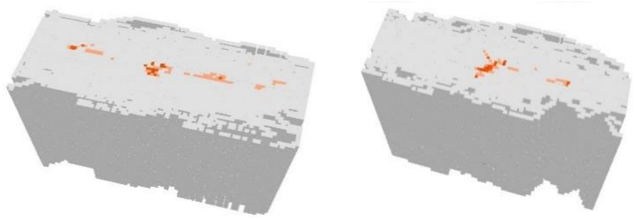

Fig. 5. The space-time cubes for both groups (left: movie with spatial sound, right: movie with non-spatial sound), there is one time coordinate and two space coordinates. For investigating the data, the slices of the cube are useable.

In this way we could see for every time interval (we used one second) where the most views are. However, the correlation between the neighbouring points (in space and time) were not considered in this step. For this we will make use of the hotspot analysis.

A problem of STC for spherical data is the distortion caused by the projection from the sphere to a plane. The further an area is away from the horizontal 0°-line the more relevant is the distortion. However, in our dataset this distortion is negligible, because nearly all data are close to the 0°-horizontal line (fig. 9). Therefore, we used the projected data also in the next step – the statistical analysis - for finding differences in the viewing behaviour watching a movie with spatial and non-spatial sound.

3.3. Getis-Ord Gi* Statistic

For finding statistically significant hotspots, methods of spatial statistics can be applied. These methods take into account the neighbouring relations between the space-time segments in the STC.

The collected data are point incident data, these are points connected to an event – in our case the viewer looked at this point. We were interested in significant clusters. To find such clusters, we used the Getis-Ord Gi* statistic [15]. This statistical method requires values for the investigated points. In order to use it, the incident data were aggregated and incident counts established. The incident counts - the number of views within a segment - are the attribute values which are analysed by the method.

The Getis-Ord Gi* statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$$

where x_j is the attribute value for point j, $w_{i,j}$ is the spatial weight between point i and j, n is equal to the total number of points and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2}$$

Using the spatial statistic tools of the ArcGIS Pro Software again, we generated STCs displaying hotspots (fig. 6).

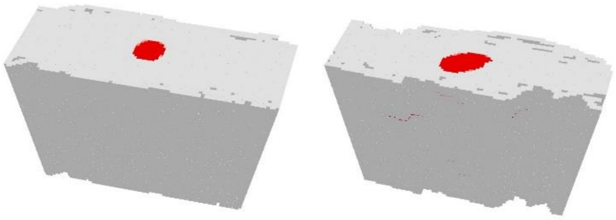

Fig. 6. The space-time cubes with significant hotspots for both groups, (left: movie with spatial sound, right: movie with non-spatial sound), there is one-time coordinate and two space coordinates.

For every point in the STC the p value can be displayed by double-clicking on the point. The p value represents the probability that the observed pattern was created randomly. A small p value means that the pattern is most likely caused by a cluster. Segments with p values smaller than 0.01, which means 99% confidence, are displayed in red.

Using the 3D mode of the GIS tool for a more detailed investigation of the data was very time consuming with our dataset. The 3-dimensional STC was suitable for an overview. However, it was difficult to navigate in the 3D cube due to the size of our dataset. We used an Acer Predator G3-710 (Intel i7-6700, 3.4GHz, 16GB RAM). Navigating in smaller datasets was not a problem. Switching to the 2D mode and modifying the time slices by a slider we could analyse the clustered areas in a comfortable way (fig. 7).

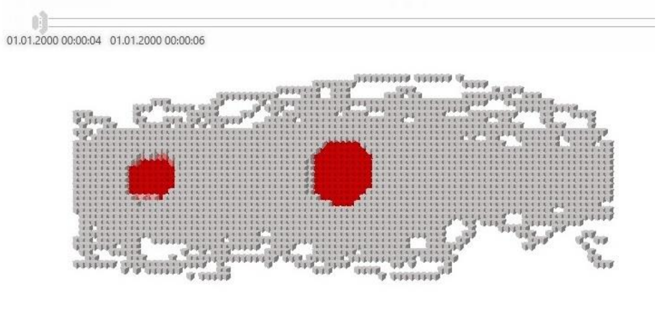

Fig. 7. Using a slider in the 2D-mode we could navigate in the STC changing the time slice. The red areas are significant hotspots.

With this approach we compared the two datasets. We changed the time slices step by step and compared the hotspots generated by using spatial sound with the hotspots generated by using non-spatial sound.

4. User Study

4.1. Material

For our study we utilized a 9:26 min movie ("Crossing Border") made by young filmmakers. This movie was produced with spatial sound and contains many cues for guiding the viewer, for example speaking persons, movements and sounds. The participants sat on a swivel chair while watching the movie via head mounted display (Samsung Gear VR with Samsung Galaxy S6) and headphones. After watching the movie, a short, unstructured interview followed to find hints if the viewer realized something regarding sound.

4.2. Participants

Our study is based on a between-subject test design. The movie was shown to two groups. One group, called group spatial here (20 participants aged between 23 and 65, mean=31.5, 10 women and 10 men), watched the movie with spatial sound, the other group, called group non-spatial here (20 participants aged between 21 and 67, mean=34.7, 8 women and 12 men), with non-spatial sound. Each group contained 6 participants watching CVR movies for the first time, and 3 experienced participants. The other participants watched CVR movies occasionally. There was no special task for the participants.

Before even starting with the session, the test persons were not told what the study is about to avoid any psychological influences with expectations of 3d audio compared to stereo audio. If we told the viewers about the comparisons we wanted to analyse, they would immediately listen more for the soundtrack than a normal consumer. Therefore, group spatial ranged from sound engineers who afterwards weren't sure if the sound was spatial, to laypersons who immediately realized that the sound field was rotating with head-movement.

4.3. Results

Analysing the tracked head movements in total, it is noticeable that in both groups the movements are very limited. We inspected the yaw and pitch angles of head movements separately. The yaw angles (turning the head to the side) were similar for both groups. Most of the time the participants looked straight forward and seldomly turned around (fig. 8). These results resemble those of [5].

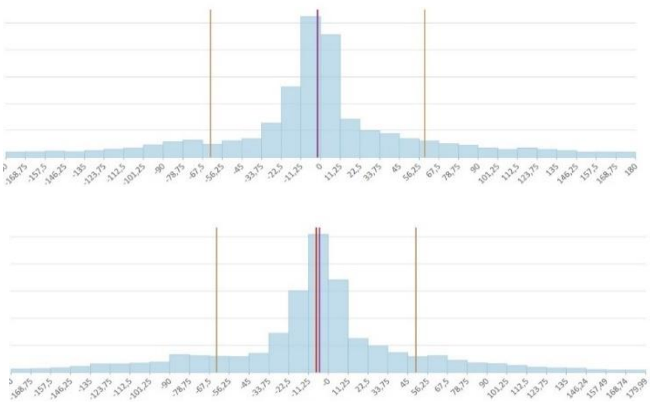

Fig. 8. The participants looked barely left and right in both groups, above: spatial sound, below: non-spatial sound (lines: red - mean, blue - median, yellow - standard deviation), the x axis shows the angle of head movements and the y axis the number of views (the views are recorded 60 per second).

Besides, participants barely moved their heads downwards or upwards (fig.9). Mostly the participants were keeping their head straight ahead, barely more than 15° downwards or 5° upwards.

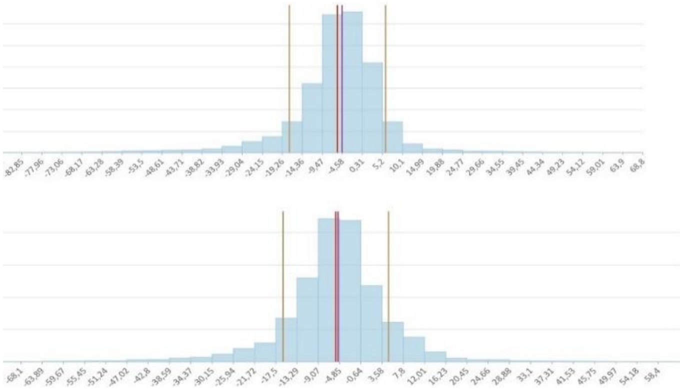

Fig. 9. The participants looked barely up and down in both groups, above: spatial sound, below: non-spatial sound (lines: red - mean, blue - median, yellow - standard deviation), the x axis shows the angle of head movement and the y axis the number of views (the views are recorded 60 per second).

One reason for that is the type of our collected data – we tracked the head movements. Near the horizontal line it is easy to move the head for changing the viewing direction. However, for looking at the ground or sky it is more comfortable to move the head just slightly and expand the viewing direction by eye movements. The sector in which a person moves the head effortlessly is around 30° in both directions [1].

Another reason for the small covered section is the setup of the movie - nearly all points of interest are nearby the horizontal line.

At first sight, the heatmaps for both groups look very similar. When persons were speaking or moving, there were areas of high data density for both groups (fig. 10).

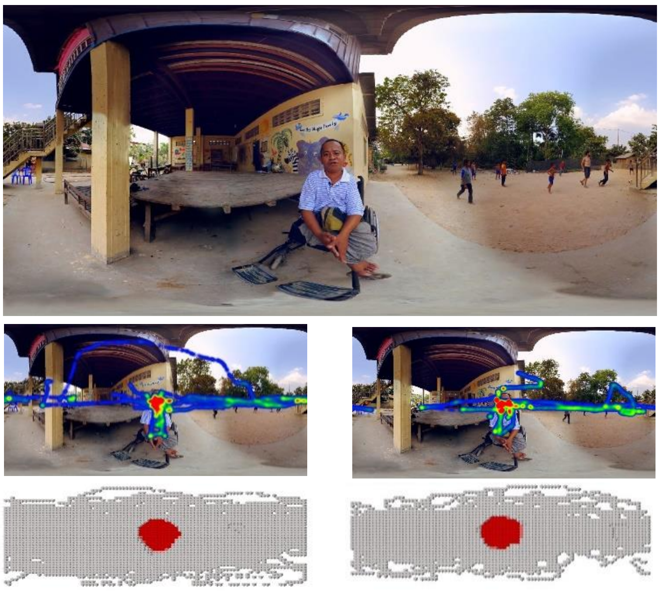

Fig. 10. Speaking or moving people attract the viewer's attention of both groups very similar (left: spatial sound, right: non-spatial sound), top: the scene of the movie, middle: heatmaps, bottom: significant hotspots.

Using the Getis-Ord Gi* statistic for identifying hotspots in the STC we found more hotspots in the data of the group which watched the movie with non-spatial sound (99 hotspots in the group spatial and 149 hotspots in the group non-spatial). Afterwards the STCs were compared, it is possible to navigate through the time slices for a closer inspection. In this way, more difference between the two groups could be found. If in a scene two or more cues occurred simultaneously (for example: speaking in one place and movements in another), the hotspots were split more often in the group that watched the movie with non-spatial sound. For example, figure 11 shows that, in addition to the hotspots around the mopeds, a hotspot around the sign has formed.

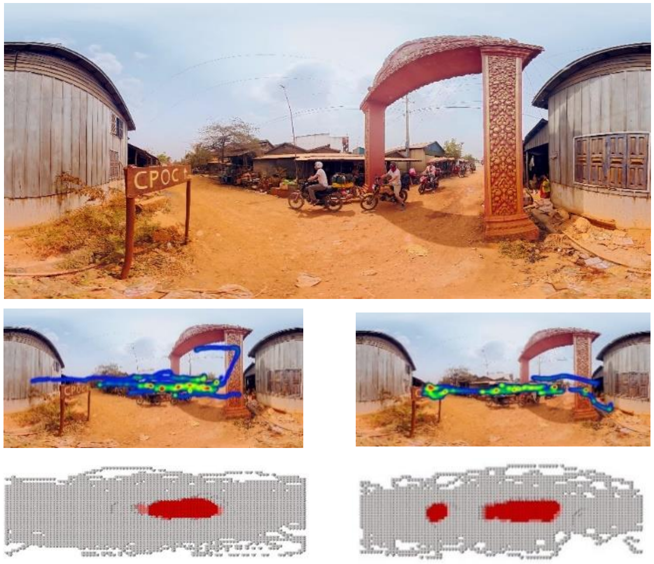

Fig. 11. Slices of the STC (left: movie with spatial sound, right: movie with non-spatial sound), top: the scene of the movie, middle: heatmaps, bottom: significant hotspots.

When people are talking to the camera, the viewer in CVR has the impression that the person is talking to him. In nearly all cases the viewers with non-spatial sound were looking to the speaker approximately 8s before starting to look around. In the case of spatial sound, it took 1-3s longer before watching for other details. In figure 12 - 15 an example is shown. The girl is talking to the camera/viewer about her life. The scene starts at 2:34 (fig.12). At 2.38 (spatial) and 2:39 (non-spatial) nearly all participants are looking at the girl (fig.13). After the girl started talking, the hotspots have not changed for 12s in both groups (fig.14). However, the participants of the group non-spatial started earlier to look around as figure 15 shows.

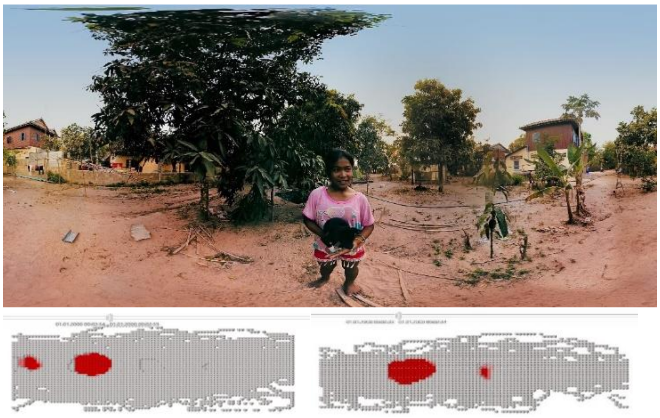

Fig. 12. Scene image and significant hotspot for time code 2:34, start of speaking (left: spatial, right: non-spatial sound).

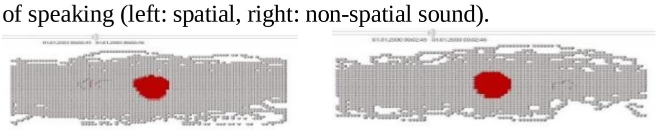

Fig. 13. Hotspot for time code 2:38 (left: spatial) and 2:39 (right: non-spatial), the associated image is nearly the same as in fig. 15.

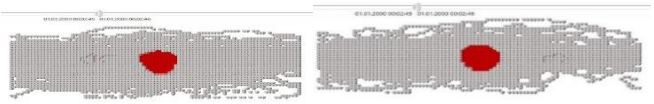

Fig. 14. Hotspot for time code 2:46, 12 s after the girl started to speak (left: spatial, right: non-spatial), the associated image is nearly the same as in fig. 15.

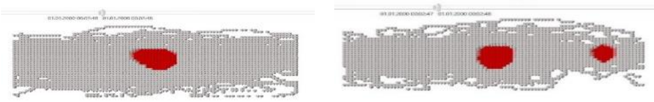

Fig. 15. Hotspot for time code 2:48, 14s after the girl started to speak (left: spatial, right: non-spatial), the associated image is nearly the same as in fig. 15.

We could find very similar results for other timecodes in the movie, where people spoke to the viewer.

5. Discussion and Limitations

5.1. Results highlighted by the Heatmap

Comparing the viewing behaviour watching CVR videos with spatial and non-spatial sound, often the behaviour was similar for both groups. Summarizing we found the following results:

- Speaking or active people draw the attention of the viewer with spatial and non-spatial sound.
- With spatial sound, the viewers observing a person stay a few seconds longer
- Details without sounds draw more attention in the case of non-spatial sound.
- If the viewers saw a scene with a setting that has already been shown earlier through the experience, such as an interview with the same protagonist in the same environment, both groups started looking away from the interviewee faster than at the first time.

We used heatmaps and spatial statistical methods for inspecting the data. Heatmaps have the advantage of being connected directly to the playing movie. So, it is very comfortable to inspect the data visually. However, the discovered results must be verified for significance using statistical methods. For finding statistically significant results we applied the Getis-Ord Gi* statistic analysing STCs. On this way we could determine significant hotspots in our dataset and compare the viewing behaviour for spatial and non-spatial sound.

The applied methods projected the movie and the data to a flat screen. This causes a distortion in areas far away from the 0°-horizontal line. In our heatmap tool this was considered, so the calculation was also correct for areas away from the horizontal line.

However, the problem of distortion was theoretically present in our spatial statistical analysis - using the STC. Since in our dataset all points are close to the horizontal 0°-line we neglected the distortion. As the STC has proved of value for inspecting tracking data, it is worth to expand this method by considering the distortion.

5.2. Results highlighted by the Survey

Since group spatial was watching talking protagonists longer, we gathered interesting opinions from the subjects why this is. One stated that she “felt rude looking away”. Like in real-life, people would look at a person who is talking the oneself. If they turned away, they wouldn’t see the fellow human being anymore, but would still hear the person talking from the side. There seems to be a relation in VR, meaning that spatial sound gives the viewers a bigger feel of presence from the virtual person that non-spatial sound is not capable of, due to the lack of localisation during head-tracking. So spatial sound seems to unconsciously remind the viewer of another person still being nearby.

Another finding can’t be highlighted by the heatmap. Some viewers of group non-spatial were confused by the voice-over since it doesn’t give an audible cue if the talking human is visible inside of the scene or not and thus a voice-over. In both cases, the dialogue was perceived mono as in head localisation and not spatial. But since group spatial could hear the interviewee spatialized with head-tracking, they immediately knew that the person was visible in the scene and a voice-over in cut-scenes if it was mono, so head-locked. This presents an intuitive way of making the viewer understand, what kind of elements are visible in the current environment and the distinction of film language.

5.3. Limitations

In our study, we tracked the head movements – not the eyes. Through that, we could follow the viewing direction in general, however not in detail. This was sufficient for a first approach and the experiments led to evaluable data. Tracking the eyes instead the head would give more detailed information but also needs more efforts for considering the distortion.

The content used in this case-study had its main action focused mainly at 0° azimuth, so there was little motivation for the viewer, to rotate more than 180° around the z-axis. But the invisible zone around the viewer is important for triggering head-movement with spatial sound and therefore might show more differences when comparing its soundtrack to a stereo-downmix.

Since the stereo-downmix represented a binaural print of the 0° position azimuth and 0° elevation, it already contained spatial information and does not represent a classic stereo mix which is also available for loudspeaker playback.

6. Conclusion

In this research we investigated where the viewer is looking in cinematic virtual reality. We compared the behaviour watching movies with spatial and non-spatial sound and found that sound draws the attention of the viewer even if it is non-spatial. These results can be used for integrating cues in a movie for guiding the attention of the viewer to things which are important for the story.

Our approach of using spatial statistical methods has proved of value and we could investigate if clusters of viewing directions are notable. This method is usable for all spatiotemporal datasets, especially head and eye tracking data. Further investigations are necessary to explore the viewers' behaviour in cinematic virtual reality.

7. References

[1] Dreyfuss, H.; others The measure of man: human factors in design; Whitney Library of Design New York, 1967;

[2] Poeschl, S.; Wall, K.; Doering, N. Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. In 2013 IEEE Virtual Reality (VR); IEEE, 2013; pp. 129–130.

[3] Larsson, P.; Västfjäll, D.; Olsson, P.; Kleiner, M. When what you see is what you hear: Auditory-visual integration and presence in virtual environments. Proc. 10th Annu. Int. Work. Presence, Oct. 25-27, 2007, Barcelona, Spain 2007, 11–18.

[4] Dreyfuss, H.; others The measure of man: human factors in design; Whitney Library of Design New York, 1967;

[5] Pausch, R.; Snoddy, J.; Taylor, R.; Watson, S.; Haseltine, E. Disney's Aladdin. In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96; ACM Press: New York, New York, USA, 1996; pp. 193–203.

[6] Brown, A.; Sheikh, A.; Evans, M.; Watson, Z. Directing attention in 360-degree video. In IBC 2016 Conference; Institution of Engineering and Technology, 2016; p. 29 (9 .)-29 (9 .).

[7] Rumiński, D. An experimental study of spatial sound usefulness in searching and navigating through AR environments. Virtual Real. 2015, 19, 223–233, doi:10.1007/s10055-015-0274-4.

[8] Gunther, R.; Kazman, R.; MacGregor, C. Using 3D sound as a navigational aid in virtual environments. Behav. Inf. Technol. 2004, 23, 435–446, doi:10.1080/01449290410001723364.

[9] Van der Burg, E.; Olivers, C. N. L.; Bronkhorst, A. W.; Theeuwes, J. Pip and pop: Nonspatial auditory signals improve spatial visual search. J. Exp. Psychol. Hum. Percept. Perform. 2008, 34, 1053–1065, doi:10.1037/0096-1523.34.5.1053.

[10] Hoeg, E. R.; Gerry, L. J.; Thomsen, L.; Nilsson, N. C.; Serafin, S. Binaural sound reduces reaction time in a virtual reality search task. In 2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE); IEEE, 2017; pp. 1–4.

[11] Hägerstraand, T. WHAT ABOUT PEOPLE IN REGIONAL SCIENCE? Pap. Reg. Sci. 2005, 24, 7–24, doi:10.1111/j.1435-5597.1970.tb01464.x.

[12] Kraak, M. J. The space - time cube revisited from a geovisualization perspective 2003, 1988–1996.

[13] Kristensson, P. O.; Dahlback, N.; Anundi, D.; Bjornstad, M.; Gillberg, H.; Haraldsson, J.; Martensson, I.; Nordvall, M.; Stahl, J. An Evaluation of Space Time Cube Representation of Spatiotemporal Patterns. IEEE Trans. Vis. Comput. Graph. 2009, 15, 696–702, doi:10.1109/TVCG.2008.194.

[14] Diansheng Guo; Jin Chen; MacEachren, A. M.; Ke Liao A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). IEEE Trans. Vis. Comput. Graph. 2006, 12, 1461–1474, doi:10.1109/TVCG.2006.84.

[15] Getis, A.; Ord, J. K. The Analysis of Spatial Association by Use of Distance Statistics. Geogr. Anal. 1992, 24, 189–206, doi:10.1111/j.1538-4632.1992.tb00261.x.

[16] Ord, J. K.; Getis, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. Geogr. Anal. 1995, 27, 286–306, doi:10.1111/j.1538-4632.1995.tb00912.x.

[17] Songchitruksa, P.; Zeng, X. Getis–Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data. Transp. Res. Rec. J. Transp. Res. Board 2010, 2165, 42–51, doi:10.3141/2165-05.

[18] Visualize Space Time Cube in 3D—Help | ArcGIS Desktop Available online: http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/visualizecube3d.htm (accessed on Jul 5, 2018).

[19] Dreyfuss, H.; others The measure of man: human factors in design; Whitney Library of Design New York, 1967;