This article has been accepted for publication in a future issue of IEEE Computer Graphics and Applications. This is a preprint. DOI: 10.1109/MCG.2021.3058129

Talking Realities: Audio Guides in Virtual Reality Visualizations

Shahid Latif, Hagen Tarner, and Fabian Beck
paluno – The Ruhr Institute for Software Technology
University of Duisburg-Essen

***Abstract*—Building upon the ideas of storytelling and explorable explanations, we introduce Talking Realities, a concept for producing data-driven interactive narratives in virtual reality. It combines an audio narrative with an immersive visualization to communicate analysis results. The narrative is automatically produced using template-based natural language generation and adapts to data and user interactions. The synchronized animation of visual elements in accordance with the audio connects the two representations. In addition, we discuss various modes of explanation ranging from fully guided tours to free exploration of the data. We demonstrate the applicability of our concept by developing a virtual reality visualization for air traffic data. Furthermore, generalizability is exhibited by sketching mock-ups for two more application scenarios in the context of information and scientific visualization.**

DATA-DRIVEN STORYTELLING is about communicating key insights gained in a data analysis to a wider audience [8]. It usually involves integrating one or more visualizations with a textual narrative. To date, the storytelling potential of visualizations has been mostly explored in the context of 2D displays. Although there are numerous approaches that focus on virtual reality storytelling in general, the research on communicating data-driven insights in a virtual reality environment is still in its infancy. On the one hand, ideas of 2D storytelling—especially with respect to what content to be communicated—are extensible to virtual reality applications. The presentation of this content in virtual reality, on the other hand, requires an altogether different approach. For instance, textual explanations—which are mostly used in 2D data-driven storytelling—are no longer a good option. Longer textual explanations are not only difficult to read in virtual reality but may also occlude the display and produce visual clutter. Besides, if not properly designed, they cause motion sickness, fatigue, and discomfort [10]. Hence, instead of text, we propose to use audio narration to communicate the story.

Audio has always been a powerful medium when it comes to storytelling. It has already been

used in various virtual reality applications such as games, movies, reconstruction of historical events, and virtual museums. Creation of such applications involves prerecording and inclusion of audio commentary either at various stages of the story (e.g., in a game) or triggered by user interactions at predefined locations (e.g., in a virtual museum). Since such applications provide a limited flexibility in terms of what users can interact with and change, the adaptability of aural content with user interactions is not a concern as everything can be scripted beforehand. In contrast, this adaptability becomes a challenge in data-driven stories where explanations need to be adapted for different datasets and user input, thereby demanding a more flexible storytelling support.

In a data-driven story, since the content is presented across multiple modalities (visual and text/audio), a consistent linking of both representations is vital. For 2D storytelling, it has already been found that the visual cues—provided as visual annotations or highlighting while listening to the related narrative—help users in bringing their attention to relevant parts of the visualization faster [4]. Accordingly, we propose to use visual annotations to synchronize the audio narration and a virtual reality visualization. The combined audio–visual representation can also reduce the split-attention effect [6]; switching between text and visualization may otherwise overload the working memory of users. Also, visual and auditory material is partly processed independently in working memory [6].

We introduce **Talking Realities**, a concept that combines a data-driven audio narrative with an immersive virtual reality visualization. The audio narrative is automatically generated and adapts to data selections and user interactions. While the narrative guides users through identified insights, the interactive visualization allows free exploration. We first describe the generic concept independent of a specific application. Then, we apply the concept to air traffic data and showcase an interactive prototype. We have chosen the air traffic data as an example due to its complex spatiotemporal nature and suitability for broad audience—nearly anyone, mostly irrespective of age and background, can understand and connect to this scenario. Afterwards, we sketch two mock-ups to demonstrate the generalizability of our concept in an information and scientific visualization scenario.

TALKING REALITIES

With our concept, we target an adequate balance between an active exploration of a data visualization and an explanation of findings through an audio narrative for providing an engaging and immersive experience. In the past decade, the paradigm of data-driven storytelling has shifted towards interactive stories that offer more and more possibilities to explore various aspects of the data. Bret Victor's *explorable explanations*[1] argues for supporting active user participation in a story. According to Victor, users should be able to use the provided environment as a testbed for critical thinking and deep understanding instead of just following the author's line of argument. The active involvement of the users, especially in a virtual reality environment, aligns with the broader definition of *immersion* presented by Isenberg et al. [3] as *"the engagement or involvement someone feels as the result of looking, exploring, or analyzing a visual data representation."* The multi-sensory representation—mapping to sensory channels such as vision and hearing among others—of data also contributes to increased immersion [3]. Similarly, Ynnerman introduces the idea of *exploranation* (a term stemming from **explor**ation and expl**anation**) [12]. It allows users to explore a visualization while guiding them to noteworthy aspects and providing explanations about the insights. Adopting these ideas, we aim at supporting varying levels of guidance and explanations during the exploration of immersive virtual reality visualizations.

We describe the generic concept (*Talking Realities*) as a kind of reusable blueprint or model that is independent of a specific application, before the next sections instantiate the concept and demonstrate its applicability in different information and scientific visualization scenarios. Our concept is applicable to virtual reality scenarios where data is represented as a single 3D visualization. We restrict the discussion to single-user applications and assume the visualization is viewed with a state-of-the-art head-mounted dis-

[1] http://worrydream.com/ExplorableExplanations

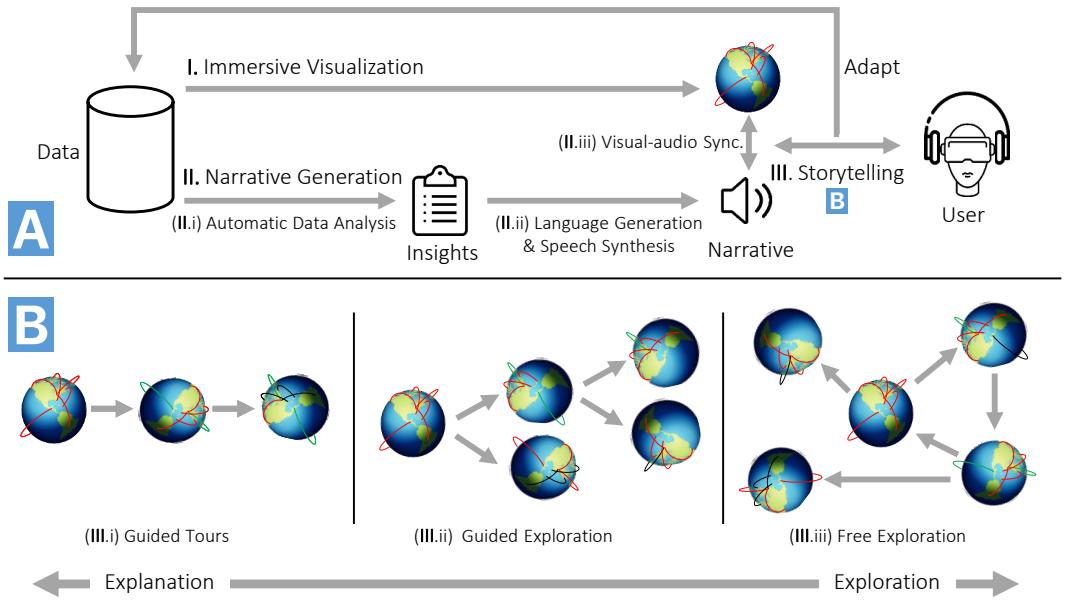

Figure 1. Abstract representation of Talking Realities—a concept for producing data-driven narratives in virtual reality. (A) The interplay of three aspects (I) *Immersive Visualization*, (II) *Narrative Generation*, and (III) *Storytelling*. (B) Storytelling offers varying levels of guidance: *Guided Tours* walk the users through a predefined sequence of events, *Guided Exploration* provides hints at various possible perspectives to explore, and *Free Exploration* enables the users to freely examine the dataset.

play. In general, we do not target an audience of experts, rather a broad group of users and do not assume specific previous knowledge about visual analysis or the data domain. For example, one such scenario is the visualization of long-distance air traffic projected onto a virtual 3D globe as it could be shown in an aircraft exhibition.

Figure 1 describes *Talking Realities* on an abstract level. The process begins with the visualization and analysis of a given dataset. While an immersive visualization provides an overview of the data and offers exploration, an automatic data analysis results in insights that are then converted to audio narrative using natural language generation and speech synthesis (Figure 1A – Narrative Generation). The story is produced by synchronously integrating the audio narrative with the immersive visualization (Figure 2). Finally, users can choose from three different levels of explanation and exploration, ranging from fully guided tours to a free exploration of the visualization (Figure 1B). We rely on existing techniques and off-the-shelf tools for accomplishing various tasks in our generation pipeline (Figure 1A).

Immersive Visualization

A visualization in virtual reality provides an overview of the data and serves as an anchor point throughout the story progression and exploration. To use the full potential of the three-dimensional virtual world, it should have a meaningful third dimension that adds value to the analysis. Such visualizations can be borrowed from existing literature on virtual reality immersive analytics for information visualizations as well as scientific visualizations [1]. To support the exploration, the *Immersive Visualization* enables interactions. We discern between two types of interactions. The first type includes zooming, rotating, and panning. Since these interactions allow users to adjust the orientation of the scene, we refer to them as *visual navigation interactions*. They are not associated with any audio guide and do not interrupt the audio playback either. This way, users may change the orientation of the visualization during a playback to get a peek from a different angle. The second type of interactions allows users to select and manipulate data items that are encoded by visual elements in the visualization. These interactions directly relate to the underlying rep-

resentation of the data and are referred to as *data interactions*. Users should be able to select data points and get details on demand, which can be presented as audio comments. It should be possible to filter or sub-select the visualized data. Generally, interactions in virtual reality are usually triggered through hand-held controllers.

Narrative Generation

Orthogonal to immersive visualization, narrative generation aims at automatically identifying interesting insights from the data by applying various analysis techniques. The resulting insights are then verbalized. The process consists of following three steps:

Automatic Data Analysis The first step is to automatically find interesting insights within the data. Table 1 provides an incomplete list of different types of analysis that can be performed. General *statistics* on the overall dataset or the most prominent parts of the data can give the users a first overview. *Clusters* of similar data items might be interesting to study as they show the main structure of the data. Specific *examples* of data items can be identified to either illustrate a representative case for a cluster or, in contrast, show noteworthy exceptions or outliers. If the users are interested in specific data items (either by selection or assumed background), a data-driven *comparison* of the items to a set of other items is relevant. Also, *trends* that describe consistent changes across sequential information such as in a time series are often of particular interest. Each type of analysis may produce a varying number of results, in total, often beyond a number that can be realistically presented to the user. In this case, we need to prioritize the findings, for instance, just listing the most prominent clusters instead of all.

Language Generation and Speech Synthesis The detected insights are then transformed to natural language text. To accomplish this, natural language generation (NLG) techniques [7] can be employed to automatically convert structured data to text. Possible options are either to use template-based text generation approaches, which work with predefined text templates, or approaches based on machine learning. Next, to convert the generated text to an audio output, any off-the-shelf speech synthesis API can be used as offered, for instance, by Google or Microsoft. These APIs allow customizing the way a text is read through the Speech Synthesis Markup Language (SSML), an XML-based markup language that controls the pronunciation and prosody of the synthesized speech. In our narrative, we discern between *contextual* and *data-driven* explanations. The former are used to introduce the dataset and provide other background information about the scenario. The data-driven explanations, on the other hand, are the ones that report notable data insights.

Table 1. Various types of analysis that can be performed for a dataset.

Analysis Type	Description
Statistics	Statistical properties that summarize (parts of) the data (e.g., average, data ranges, correlations)
Clusters	Data items of similar properties or dense connections
Examples	Single data items that are representative for a group or noteworthy outliers
Comparison	Contrasting data items to a set of other items
Trends	Changes in sequential information

Visual–Audio Synchronization Since the content is presented across two different media (audio and visual), it becomes important to temporally align these representations. Figure 2 illustrates the synchronization of content during the narration. The respective parts of the visualization are either highlighted or animated in synchronization with the related *data-driven* audio narrative (green and orange blocks). *Contextual* narrative (gray blocks) are usually not directly associated with the visualization and hence cannot be synchronized.

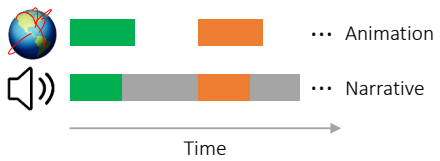

Figure 2. The visualization is animated in synchronization with the audio narrative.

Storytelling

For communicating the insights to the users, we design the progression and flow of story via interactions and target different usage strategies. On the one hand, the data-driven audio explanations provide guidance to the users while, on the other hand, the visualization allows a *Free Exploration* of the data. Considering this as a continuum between explanation and exploration (see Figure 1B), we suggest the following three specific usage modes, all three including elements of explanation and exploration but with varying strengths.

Guided Tours Focusing on *explanation*, *Guided Tours* correspond to a fully automated story including a predefined sequence of events. They are similar to self-running presentations [5]. The main objective is to present the users with a series of insights in an adequate detail. A linear sequence of available insights gets automatically selected from the data. For instance, in case of air traffic data, *Guided Tours* can walk the users through the busiest airports of the world; the airports and sequence might be different depending on the data currently loaded. These tours serve as the starting point and can be used to get familiar with the data and visualization, similar to a tutorial. Users are free to select from a list of different, but independent tours according to their interests, which provides a minimum degree of *exploration*.

Guided Exploration In contrast, the *Guided Exploration* scenario allows users to choose between various story branches at fixed junction points. The users begin by looking at the visualization and hearing an introductory audio narrative. Then, as shown in Figure 1B, at the first junction point, the users are presented with different possible story lines and can interactively choose one. As a third type of interactions, we refer to these as *story navigation interactions* since they determine the subsequent story line. The system needs to clarify the available options, for instance, using self-explanatory icon images or explaining available options as part of the audio narration. This is repeated recursively for several levels. In that way, the users navigate through the data in a mix of short *explanations* and *exploration* through choosing the next option. Different starting points should be provided for entering different *Guided Explorations*.

Free Exploration Kosara and Mackinlay [5] suggest that *"opening up a visualization for interaction at the story's end provides a convenient starting point for exploration [...]."* The *Free Exploration* enables users to further investigate details after having viewed a *Guided Tour* or *Guided Exploration*, but it can also be considered as an alternative to these modes. For instance, users may have some hypotheses in mind that they want to validate. They select data points and activate short *explanations*, which provide details on demand. Various such options for interactions should be provided so that users can steer in the desired direction. Through selecting from these various possibilities, users create their own stories fitting to their current information interests.

Both *data interactions* and *story navigation interactions* are associated with the audio narrative. Upon triggering them, the audio starts playing with the relevant parts of the visualization animated. Since the audio comments may take a while to complete, they might interfere with follow-up interactions. While *visual navigation interactions* (such as zoom, pan, rotate, etc.) should not affect the playing audio, *data* and *story navigation interactions* should influence the audio as they clearly indicate that the user is now interested in something else. When the users trigger a *data* or *story navigation* interaction while an audio comment is playing, the active playback immediately jumps to the newly triggered comment. Among other *audio controls interactions*, the users should further have the option to skip any audio playback at any time, which introduces a fourth type of interactions. Since *Guided Tours* provide limited exploration, they only cater few *story navigation interactions* for choosing or switching between the available tours. *Guided Exploration* consists of both *data* and *story navigation interactions*, where the former may also serve as an anchor point to a branching follow-up story (controlled by the latter). Finally, *Free Exploration* only provides data interactions. However, *visual navigation* and *audio controls* are available in all modes.

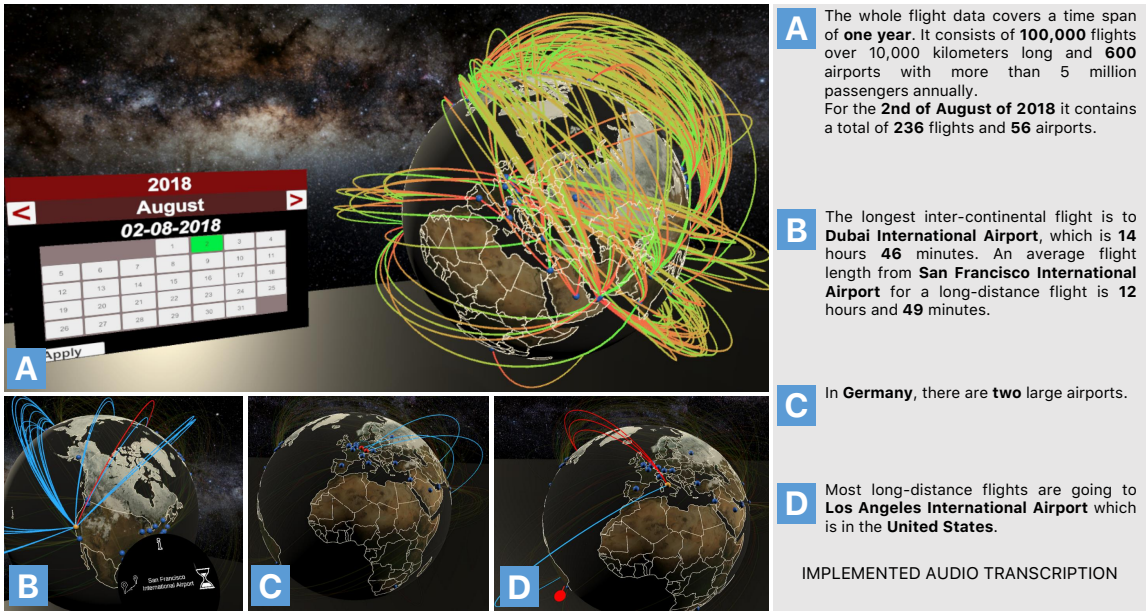

Figure 3. Graphical interface of the prototype and animations. (A) The visualization shows aggregated inter-continental flights for one regular day. The color gradient (red:departure to green:arrival) denotes the direction of flights. Animations showing (B) longest flight from an airport, (C) large airports of a country, and (D) most flights to any other airport. The right (gray) box discloses the audio transcription that are played when users see corresponding animations.

APPLICATION: AIR TRAFFIC DATA

To showcase the applicability of our concept, we apply it to a dataset containing inter-continental air traffic and describe the resulting virtual reality application. The choice of this scenario is grounded in the fact that the data has a 3-dimensional representation; visualizing it in a virtual reality environment comes naturally and makes sense. Also due to its complex geo-temporal nature, it can benefit from explicit audio explanations while showing it to the general public. Our target audience includes flight enthusiasts, school or university students, or anyone who wants to explore and get insights from global flight data. The prototype implementation uses *Unity* and *C#*; it is developed for the *Oculus Rift* head-mounted display. The dataset we use is freely available at *OpenSky Network*. It consists of more than eleven million flights for a single calendar year. Visualizing such a large amount of flight trajectories in virtual reality would result in slow rendering times and cause visual clutter. Also, short flights are not particularly relevant on a global scale. Therefore, we restrict ourselves to only large airports (visited by five million passengers annually and inter-continental flights longer than at least ten thousand kilometers. This filtering leaves us with about a hundred thousand flights for the year 2018 and 600 airports.

Immersive Visualization As we target an audience of non-experts, we went for a straightforward visualization of 3D trajectories as opposed to more complex visualization techniques [2]. The inter-continental flights are visualized on a 3D globe of the Earth as colored trajectories starting and ending at airports. Figure 3A shows an overview of all the inter-continental air traffic on August 2, 2018. We visualize airports as blue spheres at their exact locations. We approximate flight trajectories using cubic Bézier curves and visualized them as smooth curved lines. Although it may be more interesting for advanced users to see the exact flight paths, it would probably confuse novices as the exact trajectories result in more visual clutter. To prevent overlapping trajectories, we randomly exaggerate flight altitudes. A color gradient (red–green) marks the departure and arrival point of a flight. To enable exploration, the visualization offers *data* and *visual navigation*

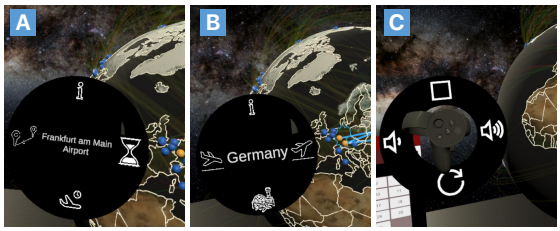

Figure 4. Virtual menus providing guidance on the possible aspects of exploration on selecting an airport (A) and a country (B). Audio controls (C) can be accessed at any time to replay, skip the current playback, and to adjust the volume.

interactions. Users can interact using the touch controllers of the *Oculus Rift*. It is possible to pan, rotate, zoom in, and zoom out (*visual navigation interactions*). For instance, users can grab the visualization and then expand or contract to zoom in or out the entire visualization. Users can interact with countries and airports to get details on demand. The calendar enables users to select and load data for different days (*data interactions*).

Narrative Generation The analysis (*Automatic Data Analysis*) results in interesting insights for all airports and countries. For each airport, we find its international as well as national (statistical) rank, number of daily departing and landing flights, longest flights to and from it, most connected airport (in terms of number of flights), and the busiest hour of the day (Table 1 – *Statistics*). Similarly, all these details—except for the last one—are detected and aggregated for each country. The analysis re-runs and updates according to data selections (for instance, when the users select another day of the year). To convert identified insights into audio guides, we first employ template-based text generation (*Language Generation*). This method of text generation works with pre-written text templates with gaps in them, which are filled with the data values. It is implemented as a decision graph consisting of pre-defined sentences for all possible cases that may arise during the generation process. For every user interaction, the decision graph is traversed from a starting node to an end node. This traversal produces a meaningful text relating to the user interaction. To ensure the correct grammar and to handle grammatical tasks (e.g., subject–verb agreement, pluralization, etc.), we leverage *SimpleNLG*—a realization engine for generating sentences from their syntactic form. Afterwards, we use a text-to-speech API[2] for transforming text into audio (*Speech Synthesis*). A combination of animations and visual highlighting helps synchronize the visual content with the audio narrative. When the audio narrative talks about a specific insight, the relevant parts of the visualization are highlighted (related airports with red and trajectories with blue color). The rest of the flights and airports are faded out using the opacity channel to create a focus–context effect. A sequence of this highlighting produces an animation effect. Figure 3 (B–D) provides examples of audio guides and the visual–audio synchronization. For instance, when the users select San Francisco International Airport (Figure 3B), the audio plays *"The busiest hour of this day is 1 o'clock, where three intercontinental flights arrive or depart."* while all these flights get highlighted (not shown in the figure); the narrative then continues to describe the longest flight from this airport as *"The longest flight is to* ***Dubai International Airport****, which is 14 hours and 46 minutes."* and the corresponding flight gets highlighted in red.

Storytelling In the system, users are confronted with the visualization and an introductory (contextual) audio that explains the application scenario and data statistics (Figure 3A). At this point, users can either go for *Guided Tours*—that are available on a virtual menu—or start exploring the visualization (*Guided* or *Free Exploration*) on their own. Guidance is provided to the users via virtual radial menus (Figure 4). These menus hint at possible aspects of data analysis that are available (*Guided Exploration*). They can be accessed through hand-held controllers. For every country and airport, detected insights are grouped into four distinct categories. For instance, when users select a country, a menu (Figure 4B) appears showing the name of the country and four possible options (*story navigation interactions*). Users can then choose among the *general statistics*,

[2] https://azure.microsoft.com/en-us/services/cognitive-services/text-to-speech

arriving and departing flights, or *large airports*. Similarly, for an airport, the radial menu (Figure 4A) contains information on *general statistics*, *temporal information* (*e.g.*, busiest hour), and the *longest flight*. *Audio controls interactions* are provided on a similar radial menu that is available on the left touch controller (Figure 4C). It can be accessed at all times and includes skip, repeat, and volume control options.

The explanations, in the form of audio guides, are presented when users interact with the globe visualization. In *Free Exploration* mode, interacting with countries or airports brings forward the details-on-demand audio guides. For instance, Figure 3C shows the result of selecting *Germany* on the globe. It highlights that there are two large airports in the country with not a lot of intercontinental traffic. Similarly, Figure 3D shows the state of visualization and audio comment while interacting with *Leonardo da Vinci (Fiumicino) Airport* airport of Rome, Italy. The comment says that this airport is most connected to *Los Angeles International Airport* via long-distance flights as three flights fly daily between the two airports.

OTHER APPLICATION EXAMPLES

Next, we sketch two mock-ups for demonstrating the generalizability of our concept. We pick two examples: one for an information visualization and the other for a scientific visualization scenario. Both examples rely on existing tools [11], [9] for producing immersive visualizations; the audio explanations, however, are not implemented and are solely drafted for mock-up purposes. While both examples begin with the mapping of our concept to the respective datasets, each one has a different focus. The first example goes deeper into the types of automatic data analysis (Table 1) and describes the *Free Exploration* mode. In contrast, the second example focuses on the *Guided Tours* and *Guided Exploration* mode. These examples may be employed in educational scenarios to explain, for instance, statistical concepts to pupils or the system of galaxies and constellations to the visitors of a virtual planetarium.

Application Example: Multivariate Data

Tabular data where objects are described along multiple variables is a common type of data. For instance, the multivariate dataset *mtcars*[3] contains eleven properties of 406 different car models that were manufactured between the years 1970–1982. It can provide insights into similar car models and relationships between different properties of cars. A scatterplot is an intuitive and established visualization that can be used for this purpose. As it is not limited to only two dimensions, more properties of cars can be simultaneously visualized as the third dimension (z-axis), color, size, shape, or opacity.

Figure 5 presents a mock-up of a possible implementation for virtual reality. At the center lies an interactive 3D scatterplot showing horsepower (x-axis), miles per gallon (y-axis), acceleration (z-axis), and the number of cylinders (color) of each car model. The visualization is produced in *Unity* with DXR [11] and offers interactions. It is possible to change (add or remove) dimensions using the menu on the left of the scatterplot as seen in Figure 5A.

Automatic Data Analysis followed by *Language Generation and Speech Synthesis* can describe insights like pairwise correlations between the selected variables (Figure 5A and C), outliers with respect to one or more variables (Figure 5B), and temporal evolution of car models with respect to the visualized variables (Figure 5D). The analysis types *Statistics*, *Examples*, and *Trends* have been used here (cf. Table 1).

To steer the users' attention, related data points on the scatterplot would be highlighted in synchronization with the audio narrative. For instance, in Figure 5A, first blue spheres and then green spheres will be highlighted sequentially (fading out the other) while the audio plays: *"Cars with a high number of* ***cylinders*** *have a higher* ***horsepower*** *but lower* ***acceleration****."*

Guidance can be provided by offering various analysis types to users, for instance, *clustering* with various types of clustering algorithms. Other users may want to directly explore the data with respect to specific questions. Figure 5 outlines an example of a *Free Exploration* scenario. It begins with getting an overview of the correlations among *horsepower*, *miles per gallon*, and *cylinders* (Figure 5A). Knowing the relationship

[3] https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html

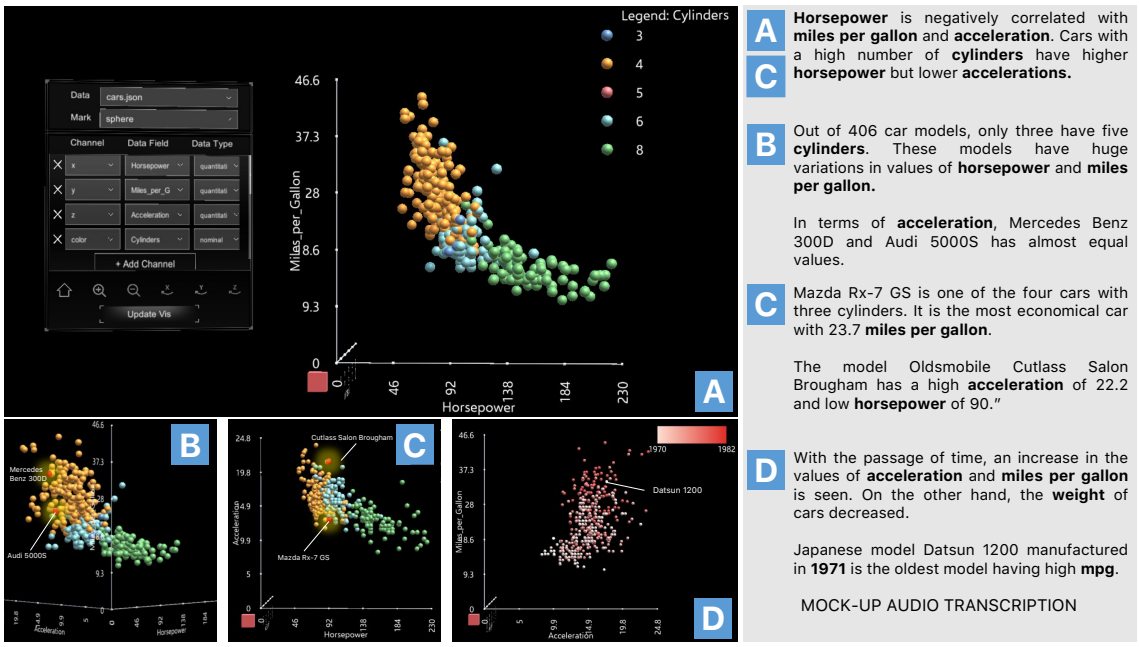

Figure 5. A mock-up illustrating the application of Talking Realities to *mtcars* dataset. (A) Scatterplot visualizes horsepower, miles per gallon, acceleration, and cylinders for 406 different car models (1970–1982). The right (gray) box shows the transcription of data-driven audio that summarizes insights related to (A,C) correlations among all visualized variables (B) unique models (outliers) with respect to number of cylinders, (C) details on demand for a selected car model, and (D) changes in the values of different properties over the years. The first text block refers to both sub-figures A and C.

between these three properties, now, the user wants to explore more the *cylinder* property; she simply chooses it from a menu on the hand-held controller. It turns out that the 5-cylinder cars are very rare and two of those (*Mercedes Benz 300D* and *Audi 5000S*) have almost equal acceleration values (Figure 5B); since one of these models is occluded by other cars, the user rotates the view to inspect it better (*visual navigation interaction*). Next, she inspects certain car models of her interest; Figure 5C describes the results of interacting with two different car models (red labeled dots). Since *Mazda Rx-7 GS* is an outlier, it has more explanation compared to *Oldsmobile Cutlass Salon Brougham*. Finally, the user decides to remove the variables *cylinders* and *horsepower*, and instead adds *year* and *weight* as variables. The *year* is added as a color gradient to the scatterplot from light red (1970) to dark red (1980) (Figure 5D). Now, the data analysis reruns and adapts to the new state of the scatterplot. The audio describes the relationship among the three visualized variables (*year*, *acceleration*, and *weight*), followed by highlighting an outlier with respect to *year* and *miles per gallon*.

Application Example: Astronomy Data

In 2013 the *European Space Agency* launched the *Gaia* Mission to create a 3D map of the Milky Way by determining position and velocity of a billion stars. From its current position at the Lagrange Point L_2 in the Sun–Earth system, the Gaia satellite records astrometric and photometric data as well as radial velocities. The current version of the dataset (Gaia DR2) contains approximately 1.7 billion objects. The observed objects include stars, planets, quasars, comets, and asteroids among others.

For visual exploration of this dataset, researchers created Gaia Sky [9]—an immersive, virtual-reality-compatible 3D visualization. We use it as our base *Immersive Visualization*. Gaia Sky offers different dimensions of exploration: observing the charted stars, exploration of planetary surfaces (elevation), and observing the satellite itself (position in orbit, movement, and

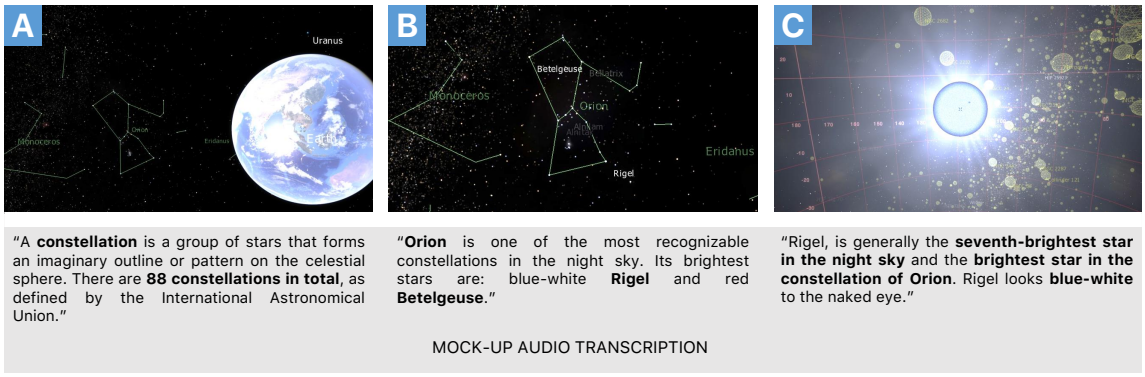

Figure 6. A mock-up of an application of *Talking Realities* to Gaia Sky: (A) an introductory text explains the concept of *constellations* to the user, (B) after selecting a constellation the generated statistics are presented, (C) the advanced users can enable optional features to display a grid for orientation (red lines) or the highlighting of nearby star clusters (yellow bounding volumes). Seen as consecutive steps, these three screenshots represent one possible story arc: starting with an overview of all possible constellations, the users first select the Orion constellation and finally analyze Rigel, a single star of the constellation, in depth.

attitude). The virtual-reality version of the application offers different controls for *visual navigation interactions*: the users can rotate the camera by moving their heads and looking around. The camera's position is changed by one of the two touch controllers. The other touch controller is used as a 3D cursor for *data interactions*: pointing at and then selecting objects. This could be extended analogous to Figure 4, to include virtual menus providing guidance on the possible aspects of exploration, like metric selection, comparison, similarity search, etc. (see example in next paragraph).

To give a concrete example of possible storytelling in Gaia Sky, the 88 star constellations designated by the *International Astronomical Union* serve as a starting point. Greeted with a welcome message (Figure 6A), the users can select one of the constellations for further analysis. The generated summary of the constellation lists its brightest stars (Figure 6B). Users can then select one of the constellation's stars. This moves the camera away from Earth and towards the selected star. Upon arrival, a summary of the selected star's metrics is played back (Figure 6C). After playback, the users can decide what to do next: follow the suggested path to the next star of the constellation (*Guided Tour*) or deviate from the path by selecting a different star or constellation for analysis (*Guided Exploration*). Navigating the story arc (or deviating from it to explore the dataset, either guided or freely) is done via the hand-held controllers. When the users select two or more stars for comparison and at least one metric, an audio description of the differences would be generated from a template, for example (with variables in **bold** font): *"Of the **two** selected stars **Rigel** is the brightest with a **Hipparcos magnitude** of **0.193**."* Filtering the amount of objects by constellation membership increases accessibility as it reduces the initial number of exploration options. On selecting a single star, a similarity search based on selected metrics could find comparable stars and highlight them in the visualization, for instance: *"**Rigel** and **Deneb** have a similar **absolute magnitude** of **-7.1**."*

FUTURE WORK AND CHALLENGES

With *Talking Realities*, we have focused on the integration of audio explanations and visually presented data. However, our scenario and solution are still limited with respect to various aspects of content presentation and human-computer interaction. These aspects can be explored as future work and include further relevant research challenges.

Evaluate the interactive storytelling: We made suggestions—as part of the abstract concept as well as specific ones in the application examples—how to design interactions for controlling the storytelling. Whether these interactions already sufficiently support the suggested explo-

ration modes and how to best achieve a smooth flow remain open research questions; extensive user testing is necessary to address this. Moreover, the storytelling can be optimized and evaluated, for instance, studying how many different story lines should be suggested depending on the previous selections of the user. Constraints are that the user should be able to advance the story at any viewing location and interactions should be effortless.

Adapt to experience levels: Considering varying levels or experience and expertise of users, also varying the level of explanation and guidance while interacting with the story would be useful. We can speculate that *Guided Tours* and *Guided Exploration* might particularly educate kids and novices in a playful manner while *Free Exploration* can be interesting to more advanced users. Still, also novices might like to play in a free manner, or the experts would profit from the hints on some advanced findings (*Guided Exploration*). Challenges include to define appropriate levels of insights and language for each group of users, to test whether certain interaction modes are preferred by different groups of users, and how a solution can automatically adapt to user experience levels.

Advance the natural feel of interactions: Users usually interact with virtual reality environments—like in the above scenarios—with motion controllers they hold in their hands. While this input mode supports basic interactions for pointing and direct manipulation of objects, other interaction modes could feel more natural and further increase immersion. Specifically, speech input would transform the audio-guide into a conversational interface, however, comes with the challenge that uses might get disappointed if the interface is not be able to answer arbitrary questions about the presented data. Also, hand gestures would be relevant for extensions of the *Talking Realities* concept. For instance, a *halt* gesture could naturally stop the current audio. In a similar way, we can think about extending the output modes, for instance, including tactile feedback synchronized to audio and visual presentation. Such feedback, however, needs to be balanced well as users might consider it as intrusive if too strong or frequent.

Leverage positional audio and avatars: Another extension to increase the immersion would be using positional audio to steer the user's attention, without force-moving the camera (which could potentially cause motion sickness). This could augment the scene with symbolic sounds (not speech) emitted from the data objects themselves. Furthermore, instead of a bodiless speaker, positional audio would also play well together with one or multiple speakers embodied as avatars in the virtual world. If having different roles, they could also help controlling a *Guided Exploration* scenario; by addressing or interacting with a certain avatar, this provides a certain context (*e.g.*, asking the *pilot* you learn about individual flights, while the *engineer* tells about the plane models in service). However, avatars, maybe more than a bodiless speaker, might come with the threat to distract and annoy the users if assisting obtrusively (compare to *Clippy* from Microsoft Office 97 and subsequent versions).

Consider multiple users and mixed reality: We have discussed the concept and application examples only in terms of single-user scenarios. However, audio explanation of *Immersive Visualizations* could be helpful also in scenarios where multiple users—in the same or at different locations—share a virtual reality experience and collaborate. New challenges arise with this more complex setting, for instance, different users might be interested in different information and hence want to hear other explanations. Another focus of scope has been on virtual reality scenarios. Augmented and mixed reality extend these scenarios with the integration of real objects, or vice versa, blending in virtual object into (a view of) reality. Interactions could be triggered and virtual visual highlighting of parts of the real object should be synchronized with the audio.

CONCLUSION

We discussed the use of audio-based explanations to guide users through *Immersive Visualizations* in virtual reality, while also letting them explore the data. As the audio comments describe the data and react to user interactions, we characterized them as adaptive data-driven audio guides. Talking Realities, the suggested concept, specifies a process of content generation, three different modes of interaction mixing *explanation*

and *exploration*, as well as the synchronized presentation of audio and visual highlighting. To demonstrate the concept as well as to illustrate its versatile applicability, we showed examples covering scenarios from geographic visualization, information visualization, and scientific visualization (one implemented in a tool, the other two sketched based on mock-ups). Whereas already practically applicable in single-user scenarios where users interact with controllers with the virtual reality environment presented on a head-mounted display, we believe to have only scratched the surface of discovering the potential of audio explanations in immersive analytics. Future promising extensions and directions for research include embedding conversational interfaces, avatars, tactile feedback, and multi-user support into the audio-guided *Immersive Visualization* scenario.

ACKNOWLEDGMENT

The authors wish to thank Arman Arzani, Bijan Shahbaz Nejad, Florian Uphoff, Jakob Robert, Nicklas Heuser, Nico Möller, Peter Roch, and Philipp Disselhoff for their support with the prototype implementation. The project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 424960846.

REFERENCES

1. T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, et al. Immersive analytics. In *2015 Big Data Visual Analytics*, pp. 1–8. IEEE, 2015. doi: 10.1109/BDVA.2015.7314296
2. C. Hurter, N. H. Riche, S. M. Drucker, M. Cordeil, R. Alligier, and R. Vuillemot. Fiberclay: Sculpting three dimensional trajectories to reveal structural insights. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):704–714, 2019.
3. P. Isenberg, B. Lee, H. Qu, and M. Cordeil. Immersive visual data stories. In *Immersive Analytics*, pp. 165–184. Springer, 2018.
4. H.-K. Kong, W. Zhu, Z. Liu, and K. Karahalios. Understanding visual cues in visualizations accompanied by audio narrations. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.
5. R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013. doi: 10.1109/MC.2013.36
6. S. Y. Mousavi, R. Low, and J. Sweller. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87(2):319–334, 1995. doi: 10.1037/0022-0663.87.2.319
7. E. Reiter and R. Dale. *Building Natural Language Generation Systems*. Cambridge University Press, 2000.
8. N. H. Riche, C. Hurter, N. Diakopoulos, and S. Carpendale. *Data-driven storytelling*. CRC Press, 2018.
9. A. Sagristà, S. Jordan, T. Müller, and F. Sadlo. Gaia Sky: Navigating the Gaia catalog. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1070–1079, 2019. doi: 10.1109/TVCG.2018.2864508
10. T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):11–11, 2011. doi: 10.1167/11.8.11
11. R. Sicat, J. Li, J. Choi, M. Cordeil, W. Jeong, B. Bach, and H. Pfister. DXR: A toolkit for building immersive data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):715–725, 2019. doi: 10.1109/TVCG.2018.2865152
12. A. Ynnerman, J. Löwgren, and L. Tibell. Exploranation: A new science communication paradigm. *IEEE Computer Graphics and Applications*, 38(3):13–20, 2018. doi: 10.1109/MCG.2018.032421649

Shahid Latif is a PhD student at the University of Duisburg-Essen, Germany. His research focuses on developing solutions for the dissemination of complex data analysis results to a wider set of audience, especially leveraging the advantages of a bi-modal representations of data containing interactively linked text and visualizations. Contact him at shahid.latif@uni-due.de.

Hagen Tarner is a PhD student at the University of Duisburg-Essen, Germany. His research focuses on Software and Performance Visualization. Contact him at hagen.tarner@paluno.uni-due.de.

Fabian Beck is an assistant professor at the University of Duisburg-Essen, Germany. He received the Dr. rer. nat. (PhD) degree in computer science from the University of Trier, Germany in 2013. His research focuses on methods for visualizing and comparing large and dynamic graphs and hierarchies, often in the context of software systems and their evolution.

He also investigates visual analytics systems and the integration of visualizations into text. Contact him at fabian.beck@paluno.uni-due.de.