Understanding Collaborative Learning of Molecular Structures in AR with Eye Tracking

Tobias Rau, Maurice Koch, Nelusa Pathmanathan, *University of Stuttgart, 70569, Stuttgart, Germany*

Tobias Isenberg, *Université Paris-Saclay, CNRS, Inria, LISN, 91400, Orsay, France*

Daniel Weiskopf, Michael Sedlmair, Andreas Köhn, Kuno Kurzhals, *University of Stuttgart, 70569, Stuttgart, Germany*

Abstract—We present an approach for on-site instruction of multiple students accompanied by gaze-based monitoring to observe patterns of visual attention during task solving. We focus on collaborative processes in augmented reality (AR) that play an essential role in on-site and remote teaching alike. From a teaching perspective, it is important in such scenarios to communicate content and tasks effectively, observe whether students understand the task, and help appropriately. In our setting, students work with head-mounted displays with eye-tracking support to collaborate in a co-located space. The supervisor can observe the scene and the students and interact with them in a hybrid setup using both AR and a desktop PC. Attention monitoring and guidance are facilitated via a bidirectional mapping between 2D structural formulas and 3D molecules. We showcase our approach with an interactive teaching scenario in which chemistry students learn aspects of stereochemistry by interacting with virtual 3D models of molecular structures. An interview with supervisors and students showed that our approach has much potential in classroom applications for (1) engaging students in collaborative task solving and (2) assisting teachers in monitoring and supporting the learning processes of their students.

Virtual (VR) and augmented reality (AR) have become available to users at home in recent years in the form of consumer-market head-mounted displays (HMDs). This availability makes AR also attractive for applications in teaching (e.g., distance education [5] or teaching for children [18]) with numerous new ways to engage with data. Such immersive technologies provide means for supplementing traditional methods with new visual and interactive ways for understanding complex educational content.

This work focuses on the subject of teaching stereochemical problems with collaborative AR (Figure 1). Learning to understand how the topology of a molecule, expressed by its abstract structural formula, maps to the actual spatial arrangement of atoms is an essential component of curricula in chemistry and related disciplines. Discussing molecules in an immersive, collaborative environment with supervisors and students has the potential to deepen students' understanding of molecular structures and stereoselective chemical reactions.

Currently, (1) physical ball-and-stick model kits, (2) interactive 3D visualization on 2D screens, and (3) VR applications are used to achieve this goal. There are benefits and limitations to all these approaches: Physical ball-and-stick models are tangible and can be explored by multiple people at a time, but they are static and cannot represent dynamic changes. In contrast, visualizations and structure modeling on a desktop PC (e.g., Avogadro [avogadro.cc], Jmol [jmol.sourceforge.net], or PyMol [pymol.org]) can represent such dynamic changes and also provide numerous different rendering techniques to emphasize specific aspects of a molecular structure [8]. Yet, the interaction with such applications is often limited to the screen-keyboard-mouse paradigm [9] and a single user.

Digital Object Identifier 10.1109/XXX.0000.0000000

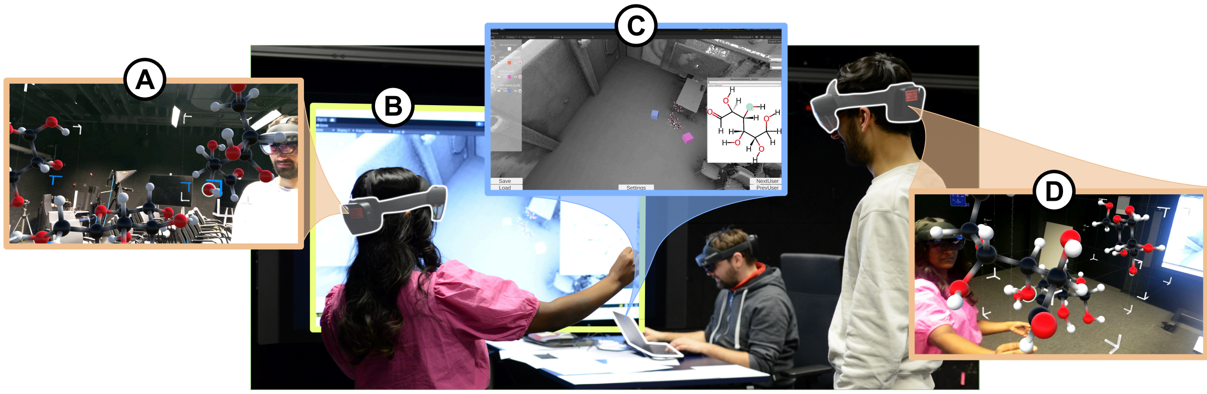

FIGURE 1. Students' perspective of the immersive environment and the teacher with the hybrid setup. Students stand and interact with the molecules in the immersive environment (A and D) while the teacher monitors the students' attention via the desktop application (C). The large screen in the background (B) was included as an optional help for the students.

VR systems (e.g., Nanome [2] or Chimera X [12]) improve the navigation in a 3D space, and the stereoscopic impression of 3D structures and are widely used by experts, pharmaceutical companies, and for education purposes. These systems have already been successfully tested for the use in VR lab courses [18]. However, they do not provide features for co-located interactions and monitoring capabilities for the teacher.

With this work, we provide a system that supports natural and flexible interaction between students, teachers, and the virtual content. We specifically take advantage of the eye-tracking hardware in the current generation of HMDs to supply the teacher with gaze-based monitoring. The gaze data is used as an approximation for visual attention, which has been investigated in the past for live interaction [13] and behavior analysis [11]. Here, we focus on the research question: *How can we support students and teachers in augmented reality by communicating the students' visual attention on the teaching content?* We expect that this approach is helpful for attention guidance, in monitoring the learning progress, for identifying problems, and providing individual feedback. In this work, we will give a first assessment in how far our new system can reach these aims.

Our implementation is based on the *chARpack* framework [15], which provides a hybrid framework for building and manipulating molecules in AR and on a desktop PC. Our contributions are: (1) We extend the hybrid interface on the desktop PC with spatial context that allows a flexible use in both spatially narrow and distributed scenarios. (2) We include a 2D–3D conversion of chemical structure formulas that facilitates interaction and attention mapping between desktop and AR. (3) By incorporating eye tracking, we support guiding and monitoring of visual attention of students. We developed segmented outlines for 3D structures and attention heatmaps for the 2D abstractions to achieve this. We demonstrate our approach with a showcase scenario based on stereochemical problems provided by our collaboration partners in theoretical chemistry, whose continuous feedback also was used to improve the implementation of the prototype. We further conducted a group interview with domain experts and students who tested the implemented system.

While we focus on teaching molecular structures due to the expertise of our co-author, our approach can also be generalized to other teaching contexts with group collaboration on virtual objects. Minor changes would also allow an application to remote scenarios instead of on-site teaching as in the present study.

IMMERSIVE VISUALIZATION AND EYE TRACKING

Representing complex 3D data can often be challenging when visualized in a 2D domain, e.g., on a regular computer monitor. Immersive visualization facilitates the visualization of such data in the physical environment using AR technologies (e.g., HMDs, tablets, smartphones, etc.). We refer to the work of Kraus et al. [6], who describe the potential of immersive visualization, particularly the intuitive way of understanding the given spatial data through interactive exploration in the 3D environment.

Collaboration becomes important when multiple persons work within the same environment and try to explain as well as jointly understand the displayed visualizations. The *Studierstube* [16] is an AR system that allows a group of users to interact and communicate simultaneously on a 3D visualization and relies on annotations to improve the understandability for collaborators. To evaluate their system, those authors studied different fields including scientific visualizations of flow and volumetric scan data, CAD models, and landscape design. While annotations can direct a collaborator's attention toward the right location, it is still important to confirm that the collaborator is focusing on the right area. It is thus essential to understand how the attention of collaborators is distributed within the environment, particularly in teaching contexts in which the teacher needs to see if students understand the given task. Numerous studies have explored the application of AR in STEM (Science, Technology, Engineering, and Mathematics) education, highlighting its potential to enhance learning experiences and outcomes [17].

Eye tracking can help analyze visual attention and understand people's task solution strategies. For the analysis of the gaze data, we can rely on visualization, e.g., in the form of heatmaps [4]. In the context of education, we can use gaze data to reason about the cognitive processes of a student while they are solving a task, as it was demonstrated in medicine [1], engineering [20], chemistry [3], and more areas. The latest-generation HMDs are equipped with eye tracking, so it is now feasible to analyze the visual attention of users for situated visualization (SV) [7]. In contrast to existing work that analyzes recorded gaze data for SV in a post-processing step [10], in this work, we contribute a real-time analysis of gaze data that helps a teacher to immediately direct the students' attention to the right place of the visual representation.

As mentioned by Bernholt et al. [3], chemistry students often struggle to understand reaction mechanisms at the molecular level when confronted with a variety of representations. It is therefore helpful if they receive hints to direct their focus toward the relevant area using highlighting or cueing based on input from a domain expert, for instance, from recordings of the expert's eye movements. Furthermore, Bernholt et al. [3] mention how VR/AR can enhance students' understanding of chemical structures, particularly when creating "3D mental models" from the 2D representations of chemical structural formulas. We combine both techniques by allowing students and teachers to see the molecular structures in the immersive environment and highlight areas of interest. In doing so, we are inspired by Wang et al. [19], who introduced a hybrid system for exploring scientific visualizations on the PC and AR and who used interaction with charts on the desktop to apply filters to the 3D AR representation. We also rely on a hybrid framework that allows us to present the 3D molecular structure both in AR and on the desktop. Furthermore, we allow the instructor to switch between the 3D molecular structure within the immersive environment and a 2D abstraction of the data with attention representation on the desktop.

COLLABORATIVE LEARNING: AN IMMERSIVE APPROACH

We propose a conceptual approach (Figure 2) to integrating teachers and students into a hybrid learning environment. This approach is implemented as a prototypical system to showcase its feasibility and benefits.

2D Abstractions

Our main idea is to provide the teacher with tools to monitor the progress and attention of the students and help them if necessary. Our precondition for this approach is a task that involves the handling of complex 3D data and that a 2D abstraction of the 3D data can be derived with a bijective mapping. Abstractions can be assembly instructions, map projections, or, as in our showcase, chemical structural formulas. Such 2D abstractions compensate for occlusion and perspective distortion problems from 3D and facilitate the interaction on a computer screen. They also play a fundamental role in chemistry, where such abstractions of molecules are the dominant representation in textbooks and scientific publications.

Visual Attention: Monitoring and Guiding

Teachers and students are familiar with the 2D and 3D representations of chemical structures, which facilitates context switches and mental mapping. In general, the 2D abstractions allow visualizing gaze on molecules as 2D heatmaps. Hence, teachers can follow the distribution of their students' attention, interpret their progress, and identify potential issues. It has the additional advantage of helping teachers monitor multiple students simultaneously, independent of their location in the room. Therefore, we provide information about visual attention in an immediate and in an aggregated format. The desktop view supports the creation of individual 2D views for each student or keeping an overview with data from all students.

For the immediate representation of gaze data, we create highlights on the 3D and 2D structures that dynamically fade according to the visual focus on the object (Figure 4). Students can see the others' gaze as highlighted segments in different colors around objects. Teachers can activate this feature to facilitate communication between students. The temporally aggregated display of gaze data is implemented with heatmaps [4]

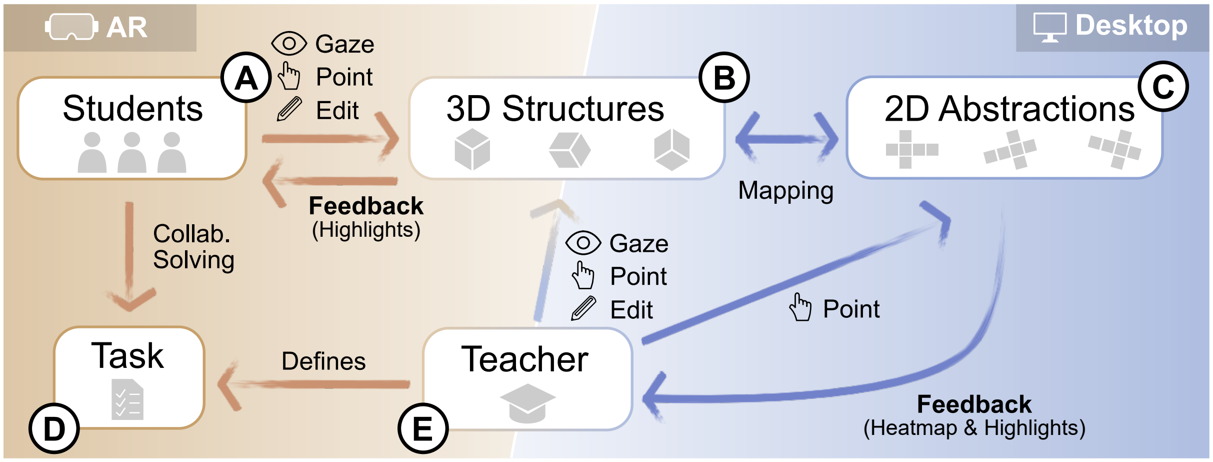

FIGURE 2. Our approach combines AR (left side) and desktop environments (right side) to provide a co-located space for immersive learning scenarios. Students (A) collaboratively solve tasks (D) by interacting with 3D structures in AR (B). The teacher (E) can interact with 3D structures in AR and its 2D abstractions (C) on the desktop. Gaze and point interactions trigger highlights on parts of the 3D structure that provide feedback to students and teachers alike. This feedback can be colored outlines on individual entities of the 3D structures. Additionally, a heatmap provides an aggregated view of the student's visual attention on the 3D structures. The 3D structures are editable, for example, by manipulating their positions or connectivity.

that indicate gaze on specific regions with a color mapping from blue to yellow with an increasing number of gaze samples. We gradually attenuate the accumulated attention of the heatmap, which is controlled by a ramp-down time.

To facilitate interaction between students and the teacher, we deploy a hybrid system in which teachers can take advantage of the bijective mapping to highlight parts of the molecule structure to guide the attention of the students—either in the 2D abstraction or directly on the 3D structure (Figure 5).

Spatial Context

To ease the switching between desktop and AR environment, we provide the spatial context of the surroundings on the desktop in the form of a point cloud. While the representation of 3D structures alone can also be achieved in VR, the interaction between students requires spatial and personal context, which is better achieved with AR. To communicate this spatial context to the teacher either in the live session or for potential later analysis, we include a digital spatial model of the working area. There are many options to derive such a model (LiDAR, photogrammetry, CAD modeling, etc.); we used a Leica BLK360 laser scanner and rendered the measured point cloud for spatial reference. For spatial alignment, we registered all HMDs with a single fixed spatial marker in the working area.

Implementation Details

Our implementation extends the framework chARpack [15], which is implemented in the game engine Unity (unity.com). The structural formula generation is done in a self-developed Python script that uses the *rdkit* library (rdkit.org). We use Microsoft's HoloLens 2 due to its eye-tracking support, but also support Meta Quest devices (2, 3, and Pro). The complete implementation of our system is freely available under github.com/KoehnLab/chARpack. Further information about the software and instructions can be found under charpack.github.io.

SHOWCASE: TEACHING MOLECULAR STRUCTURES

We demonstrate our approach with a scenario for the collaborative investigation of molecular structures. Figure 3 shows details of the internal processes. Several representation methods exist for showing 3D models of molecules. We chose the ball-and-stick model as it is easier to interact with single atoms compared to a stick-only (liquorice) representation. The ball-and-stick model is also advantageous for projecting the gaze information.

Student Perspective

In the immersive environment (Figure 4, left), students can select and manipulate molecules with established

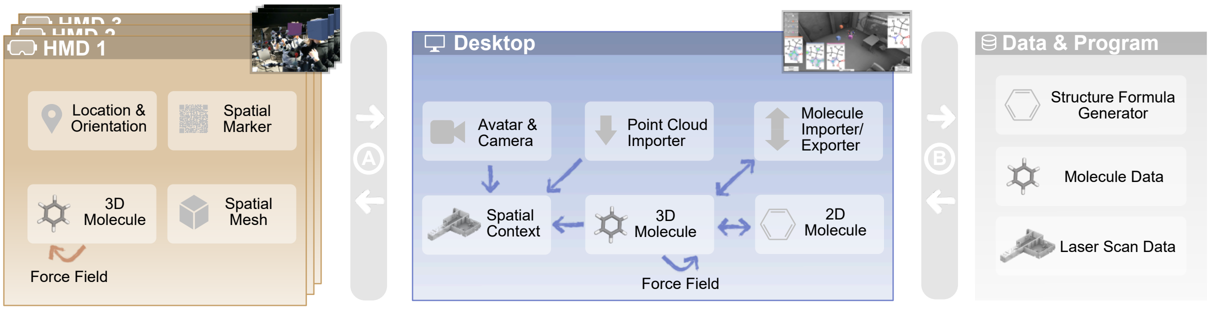

FIGURE 3. The desktop communicates with the HMDs and the structure formula generator via the network. The spatial marker is scanned during setup and ensures correct spatial alignment of clients and spatial context. Each HMD sends its current location and orientation relative to the initially scanned spatial marker to the server application (A). On the server, each HMD is rendered as a simple avatar and provides selectable camera views. We provide spatial context via the HMDs' spatial mesh or a laser scan point cloud that can be loaded by the point cloud importer via the communication interface (B). Molecules can be either built or loaded from a file (B) and are synchronized across devices. We generate the 2D structural formulas with a Python script that takes the positions and chemical abbreviations of the atoms as input.

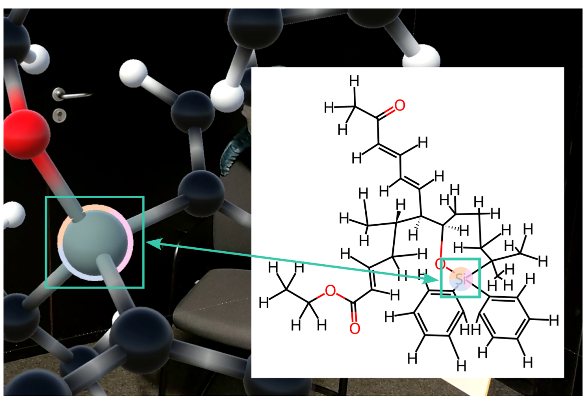

FIGURE 4. On the left, the current attention highlight is shown from the perspective of a student. A fraction of an outline is rendered around the atom. Attention is also highlighted in the chemical structural formula on the right. In both cases, longer attention to a single atom increases the alpha value of the color. On the corresponding segment of the outline of the 3D structure, the radius is additionally ramped up.

pinch and air-tap gestures. New atoms are spawned in front of the student via a selection menu. Gaze on atoms is displayed by segmented outlines, based on the number of connected HMDs (Figure 4, left). We indicate highlights by (1) the position of the outline segment around the atom, (2) the color of the outline segment, (3) the alpha value of the segment, and (4) the radius of the outline segment. For familiarity reasons, we implemented a similar highlight for the 2D abstraction (Figure 4, right), yet without the radius as a parameter. We only use this technique for immediate highlights because long-term visibility can potentially clutter the 3D structure.

2D Abstractions

We generate and synchronize the structural formulas (SVG files) with a script that connects to the desktop application via User Datagram Protocol (UDP). In addition to the SVG output for the structural formula, we provide the 2D positions of atoms for bidirectional mapping. Furthermore, we altered the appearance of the structural formulas to include hydrogen atoms, which are often omitted in standard representations but can be essential for understanding the stereochemistry. For large atoms, however, this information has to be omitted, and gaze is remapped to connected atoms.

Teacher Perspective

Teachers can guide attention by highlighting atoms in the 2D view which is linked to the 3D representation of all students (Figure 5). Also, teachers can open individual views for each student that provide heatmaps with aggregated gaze information on the structural formula (Figure 6 (E)). The heatmap is based on discrete positions on the map (atoms in our case) that can be a point of interest. Teachers with an HMD can also partake in the task with the students and investigate the task from different perspectives. In addition to providing attention guidance on the desktop, the teacher can also guide attention in the immersive environment via gaze or pointing gestures.

A Bird's Eye View

We also provide teachers with an overview of the entire scenario, using a free camera placement in the virtual scene (Figure 7 (A)), reminiscent of perspectives from real-time strategy games. This view is useful especially if multiple molecules are presented and students

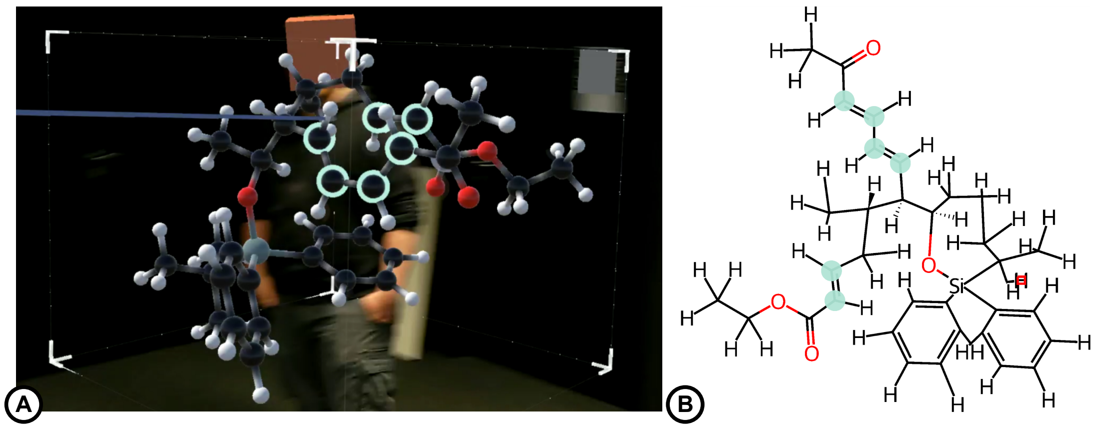

FIGURE 5. Illustration of attention guidance by the teacher. In panel A, the view of the students is shown, the highlighting appears as full circles around atoms. On the desktop application, the teacher can select atoms to be highlighted directly in the chemical structural formula (panel B). The scene illustrates task 1 of the experiment, where the atoms involved in an intramolecular reaction have to be identified. It also shows how the 2D/3D mapping can help connect topological and spatial molecular models.

distribute to different locations; then, this perspective provides the teacher with the necessary information about current spatial arrangements.

Looking Through Their Eyes

On the desktop, teachers can take the virtual perspective of individual students (Figure 7 (C)). This perspective shows the student's focus in AR (Figure 7 (B)), which supports the teacher in assisting with individual problems. While this perspective is a valuable means for instant assessment, some problems and task-solving strategies only become visible over time and require the aforementioned heatmaps for observation.

EXPERIMENT

To gather feedback from domain experts and users, we conducted an experiment in the form of a system demonstration followed by a semi-structured interview.

Participants

The experiment included two volunteering doctoral researchers in the role of teachers (both male), and six undergraduate students (3 female, 3 male) from the chemistry department of our university. All participants were introduced as a group to the system and then performed the given tasks in teams of three, consisting of one teacher and two students (see Table 1). Overall, the experiment lasted approximately 2 hours, and the students were compensated 24.81 Euro each.

Setup and Apparatus

All participants were present in the room for the whole experiment. Students who did not actively participate

TABLE 1. Distribution of the students and teachers (each numbered) for the three tasks.

task	teacher	students
1	#1	#1, #2
2α	#2	#3, #4
2β	#2	#5, #6

in a task could follow the actions from an external perspective. They watched the currently active students and the teacher on a separate large display (see Figure 1). With this setup, we ensured that all students got a good understanding of the system from different perspectives and with different tasks. We provided the teacher with a laptop on a desk, while the students were solving the task in the area in front of the desk. The content of the large screen mirrored the teacher's desktop view and was visible to all participants.

Experiment Procedure

In the first phase ($\approx$10 min), the students and the teacher calibrated the device to their eyes for gaze input. Next, students tried out the possible input methods in an individual environment without collaboration features. They learned how to build, manipulate, and analyze molecules. In parallel, we introduced the teacher to the desktop application and the possible interactions with 2D structural formulas, individual heatmaps, perspective switching, and data loading. Both teachers were already familiar with the immersive part of the application but not with the collaborative gaze features. This phase was repeated for each set of students. In the second phase ($\approx$20 min), the students connected to the collaborative environment; the teacher loaded

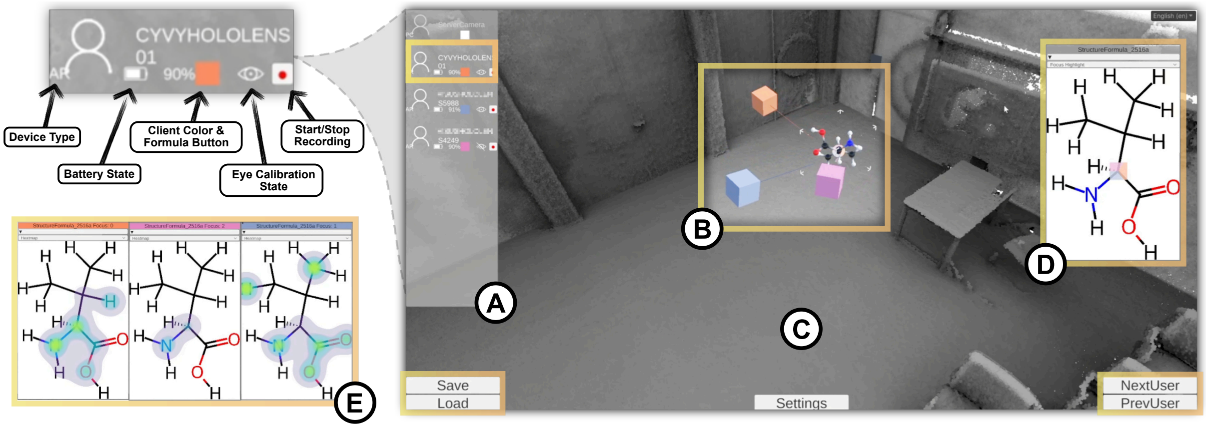

FIGURE 6. View on the desktop part of our system. The teacher is provided with a list of connected devices (A) on the left and detailed information about their state and assigned color. By default, the teacher can freely move a camera through the scene, which is also displayed in the list as *ServerCamera*. The 3D molecule and positions of the devices (B) are put into context using laser scan data that is rendered as a point cloud (C). Structural formulas of molecules that are currently present in the scene (D) automatically pop up on the screen and display the attention of all participants either using the current highlight or a heatmap. For each connected device, the teacher can open individual structural formulas that only show the attention of the corresponding participant (E), which can again use either the current attention highlights or a heatmap. Molecules can be saved and loaded into the scene using the *Save* and *Load* buttons in the left bottom corner. With the *NextUser* and *PrevUser* buttons in the right bottom corner, the teacher can cycle through all the students' perspectives and the free-to-move camera.

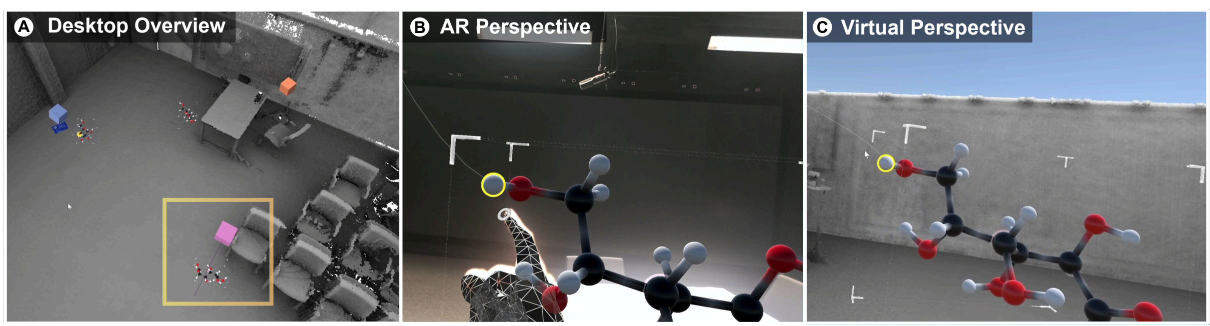

FIGURE 7. Depiction of three different perspectives. A bird's eye view shows the current position of the students and the arrangements of the molecular structures in the room (A). A part of the scene, as indicated by the colored box in (A), is shown from the students' perspective as seen from AR (B) and the desktop (C).

molecules and started to interact with the students. From the beginning, we provided the teacher with individual heatmaps of the students' attention. The second phase was repeated three times, in which the students performed different tasks. The first task was based on one given molecule structure, the second task included a variation of the underlying dataset. In the third phase ($\approx$40 min), all participants were interviewed as a group.

Task 1: Diels-Alder-Reaction

The first task involved an intramolecular Diels-Alder reaction, in which a double-ring system with a specific stereochemistry is created (Figure 5). The subtasks for the first pair of students were to (1) identify potential intramolecular reactions, (2) find the atoms involved in the reaction, (3) manipulate the molecule to create the expected product, and (4) think about alternative resulting structures of this reaction.

Task 2: Glucose

In this task, the second student pair was presented with α-glucose, while the third pair worked with β-glucose (Figure 8). In both cases the students had to (1) find existing and potential stereocenters, (2) determine the relative conformation of the hydroxyl group next to the

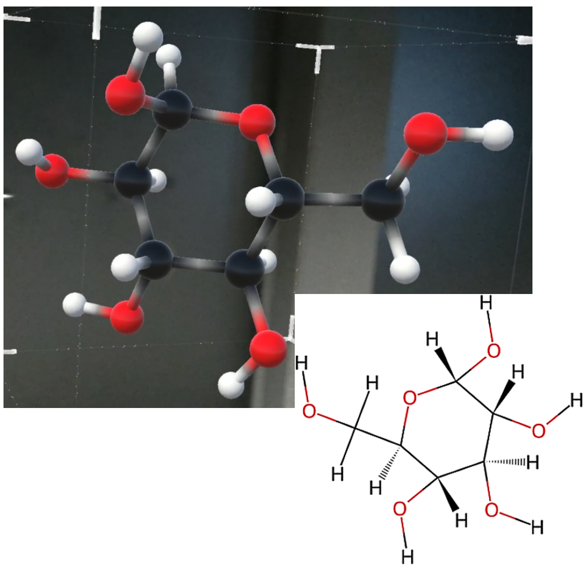

FIGURE 8. Picture with the 3D structure of α-glucose on the left and the structural formula on the right. Depending on the viewing angle, the 3D structure and the structural formula are not aligned. Hence, one of the (implicit) tasks for the students is to find a familiar viewing angle.

ring-oxygen atom and, from there, identify whether it is the α or β variant, and (3) manipulate the structure to form the open-chain aldehyde form of glucose.

RESULTS

Our results are structured into observations made during the execution of the tasks and feedback from the participants in the post-experimental interview.

Observations during Task 1

During Subtasks 1 (identify reactions) and 2 (find atoms), the teacher closely monitored the attention heatmaps of the students. The teacher provided verbal hints and highlighted atoms for students who, according to the heatmap, did not focus on the correct region. Students responded by fixating their eyes on the region of interest and with occasional pointing gestures. When tasked to find a certain double bond, the teacher detected in the heatmaps that one student searched in the wrong region of the molecule and followed up with additional hints. In Subtask 3 (manipulate molecule), the teacher moved closer and watched the students' actions in the immersive environment. The students correctly applied necessary gestures but connected the wrong atoms in their first attempt. The teacher deleted the current structure and reloaded the initial structure. For Subtask 4 (think about alternatives), the teacher switched back to the desktop application and loaded two possible products for comparison. After a brief discussion, the teacher highlighted the atoms involved in the reaction to form the final product using a combination of the 3D structure on the desktop and the 2D structural formula.

Observations during Task 2

During Subtask 1 (find stereocenters), the teacher used the individual attention heatmaps to check if all students found all stereocenters, while occasionally checking the 3D structure in the immersive environment as well. The students responded by pointing to individual atoms to react to questions and instructions from the teacher. In Subtask 2 (determine conformation), the teacher used highlighting on the 2D structural formula to help the students find the correct location of the relevant hydroxyl group. During Subtask 3 (manipulate structure), the teacher observed the students from a distance in the immersive environment. Subsequently, students #5 and #6 performed the next part of the task. For the third group of students, during Subtask 1, the teacher did not open the individual heatmaps but occasionally checked on the summarized immediate highlight in the 2D structural formula. For Subtask 2, the teacher guided the students' attention to the 2D structure on the large screen and asked them to rotate the 3D structure for better alignment. The students rotated the molecule correctly and could tell the orientation of the hydroxyl group. Subtask 3 was also observed by the teacher from a distance in the immersive environment.

Interview

We conducted a semi-structured interview with the group after the tasks. The discussion was guided by the different aspects of the system and potential applications outside of the experiment. The resulting comments are summarized in Table 2. Overall, all aspects were deemed useful, and further improvements were suggested by students and teachers.

Visual Attention *(Attention guidance, attention tracking, bidirectional mapping)*

Teachers assessed the 2D structural formulas with bidirectional mapping to be helpful to guide and track the attention of students, but noticed that some practice is necessary to deploy the technique. Students also profited from the gaze highlights. However, in the immersive view, the visual difference between gaze and selection highlights could be increased.

System *(Usability, collaboration, spatial context)*

For improved usability, the teachers requested that the global settings configuration from the desktop view

TABLE 2. Table of the interview results based on aspects considering visual attention features, usability, and application scenarios. The feedack of students and teachers is encoded by ➕ (positive), ➖ (negative), and ● (feature request).

Aspect	Teacher	Student
Attention guidance	➕ Essential when discussing structure ➕ Selection on 2D structural formula more efficient, than in immersive space	➕ Very important for communication between students ● Highlight of whole fragments
Attention tracking	➕ Heatmaps provided easy detection of students wrong attention behavior ➕ Very interesting to see different approaches to solve tasks in the heatmaps	➕ Gaze highlights did not cause cluttered view ➕ Provided means to correct wrong attention behavior ➖ Single color highlight and more prominent highlights preferred
Bidirectional mapping	➕ Bijective mapping helps locating features in 2D and 3D ➕ With preparation important regions can be located in 3D space and in 2D abstraction beforehand ➖ Practice and preparation time needed to understand mapping	-
Usability	● Configuration of global settings in immersive environment ● Manipulation of molecular structure using 2D representation	● Access to 2D structural formula in immersive environment useful
Collaboration	-	➕ Discussion and analysis beneficial on same molecular structure ➖ Difficult to work in parallel on same molecular structure
Spatial context	➕ Useful in case of seperate working areas	-
Application in lectures, exercises, tutorials	-	➕ Wish for regular use of the system in courses that require structural thinking ➕ Interaction with system leads to better understanding ➕ Loading pre-built and save structures, availability of almost infinite building material makes it superior to a traditional building kit
Further comments	-	➕ Audio feedback helps for a successful input ➕ AR approach better than VR ➖ Bonding sites hard to distinguish from hydrogen atoms ● Double bonds on 3D molecular strucutre ● Creating bonds between atoms without explicit bonding sites would simplify building and manipulation

also be included in the immersive environment. Additionally, they asked for the ability to manipulate the molecular structure using the 2D representation within the desktop view. The students requested access to the 2D structural formula in the immersive environment to better understand the 3D molecular structure. While students found it beneficial to discuss and analyze molecular structures together, they perceived collaborative manipulation as more challenging. The teachers perceived the spatial context as useful if distributed work areas were provided.

Application Scenarios *(Application in lectures)*

Students preferred the system over physical ball-and-stick model kits for its infinite amount of building material and storage of structures. They wished for timely use of the system in courses.

Further Comments

Students further preferred the AR application over VR alternatives because of the embedding into the real world. They commented that the bond representations should resemble those in common desktop tools, for instance, with double bonds. The framework chARpack uses dummy atoms as potential bonding sites. Making bonding sites more distinct from hydrogen atoms or completely replacing the need for explicit bonding sites for building was requested.

Performance

The size of the molecular system before encountering frame drops is limited by the custom force field of chARpack ($\sim$ 400 atoms). For analysis tasks (deactivated force field), the system allows the exploration of molecules with a size of approximately 2000 atoms.

DISCUSSION

In general, the system helped teachers and students collaborate closely. The presentation of the 3D molecular structure allowed the students to focus on the task, especially when the teacher was explaining specific facts by highlighting the relevant areas. This allowed them to follow the explanation, without losing the context. The teacher was able to verify whether the attention of each student was directed toward the correct location by viewing the heatmaps and could lead them toward the right area if needed. While solving the tasks, we saw that the students communicated their intentions without any difficulty. The gaze-based highlighting allowed the students to comprehend each other, without explicitly pointing at specific molecules while explaining. We found that the combination of 3D and 2D molecular representations was also appreciated by the students and helped them better understand both the topology and spatial structure of the molecules. In fact, making this connection is essential for chemists and we plan to make extended use of this feature in upcoming projects. We also noticed some technical issues with disconnects and misaligned content during the experiment. This was likely due to the resilience of the network and synchronization issues with the devices. We could solve this issue by reconnecting and resynchronizing with the server.

Lessons Learned

For future immersive learning environments, we provide a list of important aspects to consider.

- Group sizes for tasks on a single 3D structure should not exceed 3–4 students because too many parallel interactions with the same object are difficult to realize. However, teachers can potentially monitor a large number of students using attention heatmaps.
- There should be an option for individual structures to perform manipulations on. Otherwise, it could happen that one student will do the manipulations, and other students will just watch.
- Forcing teachers to switch context increases their workload unnecessarily. Therefore, we recommend supporting as many interactions as possible in both environments. Returning to the desktop for a quick settings change, even though mouse and keyboard provide better interactions, is often not beneficial.
- Spatial context is important to keep track of the physical and the virtual environment when working on the desktop. This becomes even more important when content is distributed in the working area.

In general, the virtual approach offers the advantages of saving, loading, and recording manipulations on molecular structures, which is not possible with traditional molecular model kits. However, further long-term studies will be required to see when a virtual approach can replace the established ones.

Generalizability

We believe that the presented concept can be applied to a variety of different disciplines. In architecture, for instance, students are presented with assembly tasks. These tasks can be enhanced using situated visualization and attention on parts, joints, or tools can be projected to a 2D projection of a construction plan. However, the application scenario requires 3D representations and meaningful 2D abstraction must be derivable.

Our system was developed using the game engine Unity to support other current and future HMDs with eye-tracking support. We demonstrated the system with the HoloLens 2, but adaption to other devices would require only minor changes to the system. A deployment to devices without eye tracking is also possible (e.g., we tested with the Meta Quest 3) but reduces the provided feature palette. However, porting our system to the Meta Quest Pro that supports eye tracking is a feasible alternative.

Practical Implications

With inexperienced students, teachers should plan with some additional time to introduce students to the system. A short tutorial video prior to a session could cut down the preparation time. If students used the system before and a device is already calibrated to their eyes, the setup time of the system is approximately two minutes. Setup requires the teacher to distribute the HMDs, execute the server application, and, if necessary, set up a wireless network. The students perform the eye calibration, execute the application, scan a marker for spatial alignment, and connect to the server application. The system is transportable and can be deployed to other locations without much effort. Potentially, the setup and calibration time could be reduced through spatial alignment tools that use the point cloud data generated by the HMD like Meta's physical colocation feature that implements shared spatial anchors (developers.meta.com/horizon/documentation/unity/unity-shared-spatial-anchors).

Limitations

Projecting attention explicitly on single atoms has the advantage of precise attention measurements. This level of precision, however, is not necessary when monitoring attention live. For the teacher, it is more important to see which area of a molecule is inspected by the students rather than single atoms. Hence, an approach that uses a larger area around the actual gaze ray could be beneficial. In our showcase, we focused on small to medium-sized molecules and a ball-and-stick representation, which is the usual representation when focusing on reactivity and the atomistic details of chemical compounds. For other use cases like complex biomolecules such as proteins, different representations (both 3D and 2D) have to be used and new strategies have to be found for the real-time force-field-based structural update. Our current experiment focused on qualitative feedback, to test the feasibility and identify shortcomings of the system. In the future, we plan to quantify behavior regarding attention guidance and monitoring with more controlled studies.

CONCLUSION

With our approach, students can interact with virtual molecular structures and teachers are provided with a 2D abstraction of molecules that supports them in observing and guiding visual attention. We tested our implemented system in an experiment with two experts and six students. Attention monitoring and attention guidance options were considered useful by the participants and were frequently used during the experiment. Students stated that engagement with molecules in an immersive environment helped them understand the molecular structure better and they considered the attention guidance of the teachers as important.

In the future, we plan to extend the live and post-hoc analysis features of our system. Many teaching scenarios can benefit from the participants' mobility and could be improved by AR. For instance, an extension could include different virtual learning stations distributed in a room. Teachers could then assess not only gaze patterns on individual stations but also movement patterns that might also contribute to the outcome of a learning process. Overall, we think that providing teachers with gaze information from students helps them assess learning processes, live and in retrospect. Gaze is a powerful input modality for interactive AR learning platforms in the future.

ACKNOWLEDGMENTS

This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 – 390740016, EXC 2120/1 – 390831618, Project-ID 449742818, and Project-ID 251654672 – TRR 161 (Project B01). We acknowledge the support of the Stuttgart Center for Simulation Science (SimTech).

REFERENCES

1. H. Ashraf, M. H. Sodergren, N. Merali, G. Mylonas, H. Singh, and A. Darzi; Eye-tracking technology in medical education: A systematic review. Med. Teach. 40(1): 62–69, 2018. doi: 10.1080/0142159X.2017.1391373.
2. S. Bennie et al.; A Virtual and Mixed Reality Platform for Molecular Design & Drug Discovery - Nanome Version 1.24. In: Proc. MolVa, Eurographics Association, 2023. doi: 10.2312/molva.20231114.
3. S. Bernholt, K. Broman, S. Siebert, and I. Parchmann; Digitising teaching and learning – Additional perspectives for chemistry education. Isr. J. Chem. 59(6–7): 554–564, 2019. doi: 10.1002/ijch.201800090.
4. T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl; Visualization of eye tracking data: A taxonomy and survey. Comput. Graph. Forum 36(8): 260–284, 2017. doi: 10.1111/cgf.13079.
5. M. Gattullo, E. Laviola, A. Boccaccio, A. Evangelista, M. Fiorentino, V. M. Manghisi, and A. E. Uva; Design of a mixed reality application for STEM distance education laboratories. Comput., 11(4): article no. 50, 2022. doi: 10.3390/computers11040050
6. M. Kraus, K. Klein, J. Fuchs, D. A. Keim, F. Schreiber, and M. Sedlmair; The value of immersive visualization. IEEE Comput. Graph. Appl. 41(4): 125–132, 2021. doi: 10.1109/MCG.2021.3075258.
7. K. Kurzhals, M. Becher, N. Pathmanathan, and G. Reina; Evaluating situated visualization in AR with eye tracking. In: Proc. BELIV, IEEE Computer Society, Los Alamitos, 77–84, 2022. doi: 10.1109/BELIV57783.2022.00013.
8. D. Kuťák, P.-P. Vázquez, T. Isenberg, M. Krone, M. Baaden, J. Byška, B. Kozlíková, and H. Miao; State of the art of molecular visualization in immersive virtual environments. Comput. Graph. Forum, 42(6): article no. e14738, 2023. doi:10.1111/cgf.14738
9. G. Lindgaard and S. Narasimhan; Mobile HCI: Thinking beyond the screen-keyboard-mouse interaction paradigm. Int. J. Mobile Hum. Comput. Interact. 1(3), 46–60, 2009. doi: 10.4018/jmhci.2009070105
10. S. Öney, N. Pathmanathan, M. Becher, M. Sedlmair, D. Weiskopf, and K. Kurzhals; Visual gaze labeling for augmented reality studies. Comput. Graph. Forum 42(3): 373–384, 2023. doi: 10.1111/cgf.14837.
11. N. Pathmanathan, S. Öney, M. Becher, M. Sedlmair, D. Weiskopf, and K. Kurzhals; Been there, seen that: Visualization of movement and 3D eye tracking data

from real-world environments. Comput. Graph. Forum, 42(3): 385–396, 2023. doi: 10.1111/cgf.14838

12. E. F. Pettersen et al.; UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci., Epub, 30(1):70–82, 2020. doi: 10.1002/pro.3943.
13. A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz; The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. ACM Comput. Surv. 55(3): article no. 53, 2023. doi: 10.1145/3491207
14. T. Qin, M. Cook, and M. Courtney; Exploring chemistry with wireless, PC-less portable virtual reality laboratories. J. Chem. Educ. 98(2): 521–529, 2021. doi: 10.1021/acs.jchemed.0c00954.
15. T. Rau, M. Sedlmair, and A. Köhn; chARpack: The chemistry augmented reality package. J. Chem. Inf. Model., American Chemical Society, 12: 4700–4708, 2024. doi: 10.1021/acs.jcim.4c00462.
16. D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavári, L. M. Encarnação, M. Gervautz, and W. Purgathofer; The Studierstube augmented reality project. Presence: Teleoper. Virtual Env., 11(1): 33–54, 2002. doi: 10.1162/105474602317343640.
17. M. Sirakaya and D. A. Sirakaya; Augmented reality in STEM education: A systematic review. In: Proc. Interact. Learn. Environ., 30(8): 1556–1569, 2022. doi: 10.1080/10494820.2020.1722713
18. G. Ucelli, G. Conti, R. De Amicis, and R. Servidio; Learning using augmented reality technology: Multiple means of interaction for teaching children the theory of colours. In: Proc. INTETAIN, Springer, Berlin, 193–202, 2005, doi: 10.1007/11590323_20.
19. X. Wang, L. Besançon, D. Rousseau, M. Sereno, M. Ammi, and T. Isenberg; Towards an understanding of augmented reality extensions for existing 3D data analysis tools. In: Proc. CHI, ACM, New York, article no. 528, 2020. doi: 10.1145/3313831.3376657.
20. T.-K. Wang, J. Huang, P.-C. Liao, and Y. Piao; Does augmented reality effectively foster visual learning process in construction? An eye-tracking study in steel installation. Adv. Civ. Eng., article no. 2472167, 2018. doi: 10.1155/2018/2472167.

Tobias Rau is a researcher at the University of Stuttgart, Germany. He is affiliated with both the Visualization Institute of the University of Stuttgart and the Institute for Theoretical Chemistry. His research interests focus on immersive analytics and novel interaction techniques applied to a chemical context. Contact him at rau@theochem.uni-stuttgart.de.

Maurice Koch is a third-year doctoral researcher at the University of Stuttgart, Germany. His research interests include eye tracking, visual analytics, and human-computer interaction. Contact him at maurice.koch@visus.uni-stuttgart.de.

Nelusa Pathmanathan is a doctoral researcher at the University of Stuttgart, Germany. Her research focuses on the visualization of spatio-temporal data recorded in 3D environments, situated visualization, eye-tracking, and human-computer interaction. Contact her at nelusa.pathmanathan@visus.uni-stuttgart.de.

Tobias Isenberg is a senior research scientist with Inria and Université Paris-Saclay, France. His research interests are interactive scientific visualization and illustrative rendering, using methods from computer graphics, non-photorealistic rendering (NPR), and human-computer interaction. Contact him at tobias.isenberg@inria.fr.

Daniel Weiskopf is a professor at the University of Stuttgart, Germany. His research interests include visualization, visual analytics, eye tracking, human-computer interaction, XR, and computer graphics. Contact him at daniel.weiskopf@visus.uni-stuttgart.de.

Michael Sedlmair is a professor at the University of Stuttgart, Germany. His research interests focus on immersive analytics and situated visualization, novel interaction technologies, visual and interactive machine learning, perceptual modeling for visualization, as well as the methodological and theoretical foundations underlying them. Contact him at michael.sedlmair@visus.uni-stuttgart.de.

Andreas Köhn is a professor at the University of Stuttgart, Germany. His research interests are electronic-structure theory, computational chemistry, and interactive visualization in chemistry. Contact him at koehn@theochem.uni-stuttgart.de.

Kuno Kurzhals is a junior research group lead at the University of Stuttgart, Germany. His research interests include the visualization of spatio-temporal data from sources such as eye tracking, video, and augmented reality. Contact him at kuno.kurzhals@visus.uni-stuttgart.de.