

What Do You Do When Two Hands Are Not Enough?

Interactive Selection of Bonds Between Pairs of Tangible Molecules

Patrick Maier, Marcus Tönnis, Gudrun Klinker *

Fachgebiet Augmented Reality (FAR)

Technische Universität München, Fakultät für Informatik
Boltzmannstraße 3, 85748 Garching b. München, Germany

Alexander Raith, Markus Drees, Fritz Kühn †

Fachgebiet Molecular Catalysis (MolCat)

Technische Universität München, Chemistry Department
Lichtenbergstr. 4, 85748 Garching b. München, Germany

ABSTRACT

For molecular modeling, chemical structures have to be understood and imagined both in their three-dimensional spatial extent and in their dynamic behavior. We have developed an AR-based system for tangible interaction with molecules using optical markers. When users bring several molecules close to one another, potential bonds are shown and the molecules dynamically change their 3D structure according to potential chemical reactions. A problem arises when users also need to select one such bond from of a multitude of potential bonds while already using both hands to manipulate the molecules. We present two gesture-based techniques, shake-based and proximity-based to solve this problem. We report on user tests evaluating these techniques with respect to speed, precision and user acceptance.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies, Interaction Styles; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques; J.2 [Computer Applications]: Physical Sciences and Engineering—Chemistry; K.3.1 [Computers and Education]: Computer Uses in Education

1 INTRODUCTION

1.1 Motivation

Molecular modeling requires very sophisticated and experienced understanding of the dynamics of the underlying chemical processes. Molecular structures have to be understood and imagined both in their three-dimensional spatial extent and in their dynamic behavior. When planning for chemical reactions between molecules (e.g. when designing a catalyst), chemists have to understand whether the desired result is sterically achievable, i.e. whether there is sufficient space for the molecules to form bonds between targeted sets of atoms on each side. In this respect, it is not enough to consider molecules to be rigid 3D structures. Rather, the forces between atoms need to be taken into account, thereby requiring a complex understanding of the dynamic behavior of all atoms involved in a reaction. Angular relationships between atoms within a molecule are no longer static, but rather depend on the impinging force fields from neighboring atoms from the same molecule, as well as from other molecules during a chemical reaction.

A large number of tools have been developed to help chemists and students visualize molecular structures. We all know the stick-and-ball models that chemistry teachers bring to chemistry classes to help students gain a basic understanding. They are very intuitive because they are real: students can touch and manipulate

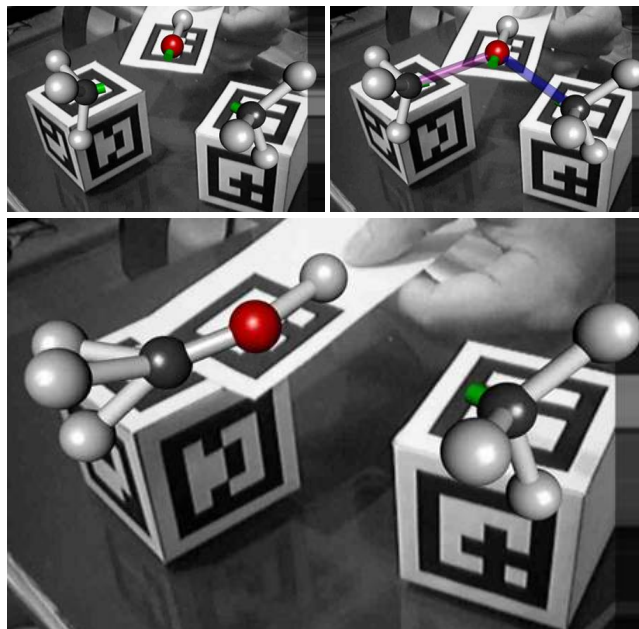


Figure 1: AR-based visualization of a dynamic adaptation of a molecule structure to forces from atoms of other molecules.

them. They can move them around and view them from all sides. However, such stick-and-ball models are inflexible. They are unable to change the angular relationships of the bonds within a molecule when new bonds occur during a chemical reaction. In contrast, various computer-based chemical simulation and visualization tools (e.g., TINKER [19], HyperChem [13]), GaussView [8] and Jmol [17]) are able to support chemists in mentally enacting and understanding chemical reactions while taking the influence of inter-atomic forces into account. Yet, the manipulation of the molecules is generally tedious. The results are typically shown as 2D or 3D visualizations. Creation of and interaction with the individual molecules occurs via WIMP-based user interfaces: users use menus, scroll bars and direct 2D manipulation to select molecules, atoms and potential bonds. In comparison to real stick-and-ball models, these simulated visualizations are harder to manipulate since they don't provide 3D handles.

Augmented Reality can come to the rescue, helping users visualize tangible stick-and-ball models on pre-defined markers, and allowing them to visualize a linear approximation of chemical reactions depending on the proximity of the controlled molecules (see Fig. 1 and Sec. 1.2). Users can interactively explore the influence of different molecular properties on the spacial representation, such as the proximity between molecules/atoms, the rigidity of the molecules as well as steric clashes, resulting, for example, in chiralic biases. By manipulating the tangible objects, users can

*e-mail: maierp@in.tum.de, toennis@in.tum.de, klinker@in.tum.de

†e-mail: raith@ch.tum.de, drees@ch.tum.de, kuehn@ch.tum.de

thus steer the simulation process. In such “hands-on” explorations and visualizations of chemical behavior, 3D positioning and timing of user gestures are an essential part of the simulation results. Users need to move molecules to the right place and keep them there while the next steps of a reaction take place. Depending on the number of molecules involved, this may require one or more hands or support structures to hold the tangibles – possibly even a team of researchers, analog to puppetry.

Now the following problem arises: Even though the direct manipulation via tangibles can control some very important aspects of a chemical simulation, many more parameters exist – and even those that are under direct positional control may be susceptible to imprecise user gestures. For example, many bonds between two molecules may be theoretically possible (see Fig. 2). If only one is to be selected, which one will it be? If it is selected based on the distance between the atoms involved (Fig. 1), this may require users to have a high level of dexterity and the ability to hold very still when non-trivial molecules are involved.

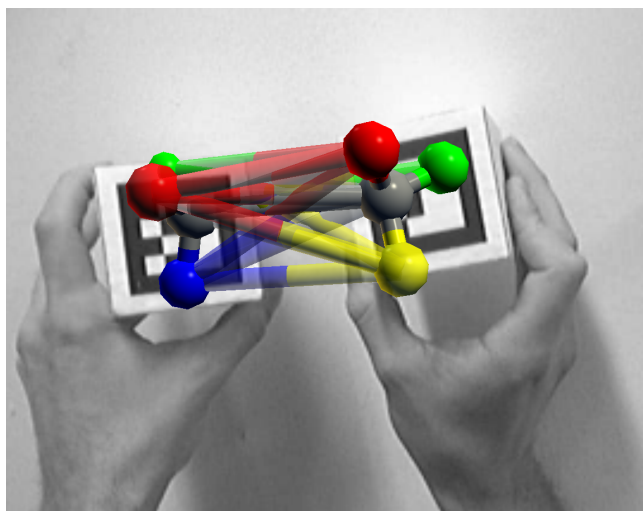


Figure 2: Exemplary display of all potential bonds between pairs of atoms on two molecules.

A number of solutions are possible to provide such system control commands alongside with direct manipulation. Some will be presented and discussed in section 1.3. We here provide a short overview of different options.

If one hand can be freed from direct tangible interaction with the molecules, it can be used to control the system via regular mouse or keyboard. It can also be used to interact with widgets that are embedded into the 3D environment [15, 21].

If, however, both hands are regularly involved in the direct manipulation of the molecules, system control gestures must be integrated more deeply. So, *what do you do when two hands are not enough?* To this end, tangibles may be equipped with extra sensors or special buttons – requiring a specialized environmental setting. If no such scene modifications are acceptable, interactive (un)clutching metaphors must be provided that allow users to switch between gestures for direct manipulation (in the true sense of Augmented Reality) and gestures for system control while temporarily “freezing” the manipulated objects. One possible option is using gestures at very different speeds: slow motions for direct manipulation and fast (meta) motions for system control or to unclutch a hand from direct manipulation. As a result of such unclutching, a hand may be freed up temporarily to perform control tasks – using any of the mechanisms discussed above.

Voice-based input or the use of foot pedals are further interactive

options. Yet, it may be hard for users to describe a specific bond concisely with spoken words or with their feet.

1.2 The “Tangible Chemical Reactions” Project at the Technische Universität München

In an on-going project at the FAR-lab and the MolCat-lab of Technische Universität München¹, we are developing a system that allows for AR-based tangible interaction to support chemists in designing catalysts for metal organic reactions and to help them understand the complex geometric and functional relationships between a suitable catalyst, the central metal and a ligand. Especially for chiral catalysis this will be of utmost importance. Fig. 3 shows an example of a tangible model of an organomolybdenum complex, used for catalysis experiments and two educt molecules, which are intended to react with the aid of the catalyst.

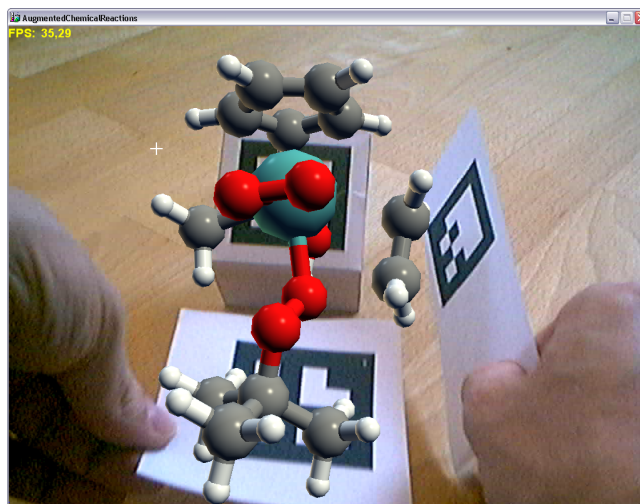


Figure 3: Tangible model of an organomolybdenum complex on the cube marker, a peroxide molecule as an educt on the front marker, making a bond with the central metal of the catalyst and a ethen molecule as the second educt on the third marker to the right.

The system presents the molecules attached to optical markers. When users bring several such molecules in close vicinity to one another, potential bonds are shown (Fig. 1). Furthermore, the molecules dynamically change their 3D structure according to the minimized energy level for a given distance and placement between the two potentially bonded atoms, defined by the user controlled physical handles [16].

The system is intended to be used very flexibly for ad-hoc investigations in MolCat labs and offices and for demonstrations in lectures world-wide. Thus, it is required to get by without special equipment beyond a laptop, a web cam and paper markers. Furthermore, users will need two hands to manipulate the molecules. Thus, no hand can be expected to be regularly free for control tasks.

We are investigating various options for providing users simultaneously with mechanisms to directly manipulate the molecules and to select specific bonds – either via a proximity-based method or by explicitly clutching between manipulation gestures and bond selection gestures. To this end, we have implemented a shake-based method to toggle through a list of potential bonds.

In this paper, we report on first results comparing the shake-based gesture with a selection method based purely on the proximity of bonding partners.

¹ www.igsse.tum.de/project/augmented-chemistry/about.html

1.3 Related Work

1.3.1 Existing Systems for Augmented Chemistry

Gillet et al. have developed a Molecular Biology application that combines the use of autofabricated (3D-printed) tangible models of biological molecules with AR-toolkit-based [14] augmented reality [9]. Their system overlays tracked physical models with different pre-recorded molecular structure representations or with textual chemical information. The structural representations are rigid, i.e.: they do not change shape when molecules approach each other. A basic animation facility is provided. The authors do not report how such animation is started or stopped.

Fjeld et al. have created an AR-toolkit-based chemical education system for children in secondary school [7]. It uses a number of specialized tangible and augmented objects to create and incrementally extend a molecule: an augmented magic book with each page presenting a different chemical element in 3D, special markers to modify the visualization and interaction mode, and a tracked gripper with a button to grab augmented elements from the book and add them to the molecule. When the user pushes the button, a grabbed new element bonds with the closest binding place on the molecule. The molecule floats in mid-air above a workspace that is defined by a movable marker on the table. The molecule can be rotated by another tangible object, a cube. Thus, users' hands are not directly attached to the molecule. Most of the time, they are free to bring in or manipulate any of a number of markers, and also to use the keyboard and the mouse to issue system control commands. In a recent publication [6], Fjeld et al. reported that users found it hard to mentally switch back and forth between tangible interactions in the augmented workplace and typing on the keyboard. In a system redesign, they have replaced keyboard entries by mouse-controlled GUIs that are embedded into the augmented environment.

Our system has a different focus than Fjeld's system. We assume that two or more tracked molecules need to be held in special 3D poses during an ongoing chemical reaction, and that their proximity has an impact on the augmented visualizations. It will be interesting to further explore in future work possible mergers and transitions between interactive metaphors of both approaches to clutch or unclutch users' hands from the molecules.

1.3.2 Approaches towards Two-handed Interaction

Two-handed interaction has so far been explored regarding two classes of concepts. The first class distinguishes between the roles, each hand has in a task, either one hand is the dominant hand and the other is non-dominant or both hands have equal roles. The second class addresses the dimensionality of the space available for task execution, either on 2D surfaces or in a 3D space.

Concerning the differences between the roles of the two hands, Guiard developed a model [10] for the asymmetric division of labor in bimanual actions. In his *kinematic chain model*, the non-dominant hand is used to coarsely define a spatial reference frame, followed by actions of the dominant hand within this reference frame at higher precision. This model has been validated, among various others, by Hinckley et al. [12] and by Xia et al. [24]. Balakrishnan et al. [1] in contrast determined deviations to Guiard's theory. They built a system for digital tape drawing on a vertical 2D surface. In contrast to Guiard's model, the designers used their right hand to define the frame of reference. Also the non-dominant left hand operated at a higher spatial frequency than the right hand. Balakrishnan et al. point out, that more analysis and refinements are required to adequately explain human bimanual interaction" [1].

Our setup uses symmetric bimanual interaction rather than an asymmetric setting. Both hands have equal functionality: 3D manipulation of the position and orientation of two molecular structures with respect to one another. Casalta et al. [5] investigated differences between asymmetric and symmetric division of labor in bimanual interaction. They set up a 2D rectangle editing task

and found that their test participants „revealed better performances and a higher degree of bimanual parallelism with the symmetrical than asymmetrical option“ [5]. They leave the question whether the Guiard's model still holds for symmetric interaction. Balakrishnan and Hinckley also investigated symmetric bimanual interaction [2]. Test users had to track a pair of targets, each controlled with one hand while forcing them to divide attention between the two targets through putting them further apart. They also investigated visual connections between the two targets. Concerning our setup, these lines can be compared to the possible bonds between two molecule structures. They found that the degree of parallelism is affected by distance and by visual cues.

The systems presented thus far use 2D interaction surfaces. Other systems offer a 3D space for bimanual interaction. The work of Pierce et al. [18] uses asymmetric two handed interaction in a virtual environment. The „Voodoo Dolls“ system provides facilities to concurrently control the working context of a handled object and its parameters for object manipulation. Grabbed by the left hand, an object is seamlessly scaled to a useful size for operation. The viewing context is adjusted according to how the two hands are held relative to each other when two objects are held. They founded their setup on the work of Guiard [10]. Evaluations showed that after a phase of familiarization, the test participants had little to no difficulties to arrange objects in a room.

1.3.3 Approaches Using Fast User Interaction (Shaking)

In our system, we distinguish between slow and fast user hand motions. Slow motions are attributed to direct molecule manipulation whereas fast motions are used as meta motions to unclutch a hand from a molecule, i.e., „freezing“ the tracking operation. Such unclutching has been used in a number of AR applications [11]. In the context of this paper, it is more important how such unclutching is achieved than what it is used for. In our case, no hands are free to push a „freeze“ button.

We have played with a bug/feature in many optical marker tracking algorithms: when the tracker loses track of a marker, the virtual object remains at the last known position. Thus, by quickly hiding a marker, a molecule can be suspended somewhere in mid-air. The freed hand can then be used for other interactions that are not linked to the molecule. However, it is doubtful whether ordinary users can be expected to be aware of the shortcomings of a tracking algorithm when using the system.

Instead, we report in this paper on a shake-based method to allow users to temporarily perform actions on a system control level. Shaking gestures have already been used in other AR-based contexts, most recently by White et al. [23, 22] to activate and deactivate menus. In contrast to our shake-based toggling through several options, White et al. use shaking only to pop up a menu. Item selection is not performed by shaking but rather by targeted marker alignment with a menu item – something we cannot do since the primary purpose of our markers is the manipulation of the molecules as a whole. Shaking is also becoming a common gesture on mobile phones with built in accelerometers [3, 20]. This indicates that our toggle-oriented approach may be a suitable addition to a growing list of use cases.

When implementing a shaking gesture, we first have to analyze how such a movement of the controlled object (i.e., the marker) can be described. White et al. have presented at technique for recognizing a shake gesture by following the path of the tracked marker and transforming it into as sequence of directional units (up, down, left, right, front, back) [22]. They parse the directional information. When they detect four continuous movements in opposing directions, they accept a shaking gesture. In contrast, we use a statistically-based approach (see Section 3.1.1).

2 SYSTEM

We now briefly describe the system setup before diving into a discussion of alternatives to select chemical bonds in section 3.

2.1 System Architecture

The system is composed of three major components: Tracking, Visualization, and Simulation. In order to be able to flexibly arrange for different technical options independently for each component, the system architecture very strictly separates these components from each other. Each component is run in an independent thread.

2.2 Visualization

The Visualization component is the core module of the system, providing the program logic, chemical knowledge and the user interface. It receives tracking information from the Tracking module. It sends positional information to the Simulation module to reflect the dynamic molecule behavior, and it receives the updated optimized molecular structure in return. It uses the DirectX-based gaming framework XNA to render molecules at the tracked locations according to the simulation state and to potential bonds between atoms of different molecules. The module also uses the tracking data to interpret gestures, as described in section 3.

2.3 Tracking

The Tracking component determines the user’s viewpoint, as well as the pose of a flexible number of tangible objects. Currently, due to the requirements of the current application scenario (section 1.2), we mainly use marker-based optical tracking, comparable in quality to the AR toolkit [14]. Poses of all objects relative to the user’s viewpoint are provided to the Visualization component.

2.4 Simulation

The Simulation module has a description of each molecule and determines its current spatial structure depending on the proximity between all modeled atoms in this and neighboring molecules. We currently use the Optimize program from the TINKER Molecular Modeling Package [19] that is based on the simulation of force fields with mass-spring models (MM3). During each optimization step, a designated atom of each molecule is fixed to the location of the marker. The remaining atoms are rearranged in the surrounding area such that the overall energy level is minimized.

2.5 Physical Layout of the Setup

We have experimented with different physical setups of our AR-chemistry environment. In our current solution, a camera is placed above and slightly behind the user’s head, looking diagonally down and forward at the user’s hands (see Fig. 7). It generates video streams from an approximately ego-centric view onto a small work area in front of the user. In this area, the user manipulates cubes with markers on each side, as well as small, flat sheets of papers with a single marker each. The camera images with the augmentations are shown on a regular monitor in front of the user’s hands.

3 TANGIBLE AUGMENTED CHEMICAL REACTIONS

When, in addition to visualizing and controlling the simulated interaction between molecules (see Fig. 1), users have to connect the molecules by selecting bonds between specific atoms from both molecules, they need to see such potential bonds, when molecules get closer than a certain distance². Furthermore, they need a method to select such a bond from a potentially large set of options (see Fig. 2). The process consists of two steps: first the preferred bond must be specified, and then the specification must be confirmed.

²The current distance threshold is 5 cm in the workspace, depending on the zoom factor for visualizing molecular structure

3.1 Specification of the Preferred Bond

In general terms, when asking users to select a bond, they are asked to specify a tuple, $bond = (atom_1, atom_2)$, from the set of tuples $B = A_1 \times A_2$, with $bond \in B$, $atom_1 \in A_1$ and $atom_2 \in A_2$. Such a tuple can be selected either by selecting $atom_1$ and $atom_2$, or by directly selecting $bond$. We will now present gesture-based methods for both of these approaches.

3.1.1 Selecting an Atom from a Set of Atoms

There are many ways to select an object a (such as an atom) from a large, composite object A (such as a molecule). In AR and VR, this has traditionally been done by pointing at the object with a tracked pointing device (or finger). Yet, in our system, the user would have to release the composite object (molecule) in order to grab the pointer. When releasing the molecule, it changes its position, which could cause unwanted effects on the chemical simulation.

Selection by Enumerative Toggling: In an alternative approach that gets by without a special pointing device, the system can ask users to toggle through a list of all objects in sequential order. To this end, a toggling signal is needed. This can be done by voice or by gesture.

In this paper, we report on a shake-based method to toggle between objects. When a user has to select one of several atoms of a molecule that have open bonds, these atoms are highlighted one at a time in a sequential order. The user toggles from the currently highlighted atom to the next one by briefly shaking the marker that is associated with the molecule. The system then highlights the next atom and continues through the ring list with every further shake by the user.

The underlying expectation is that a brief, sharp shaking gesture is less disruptive to the overall chemical simulation than depositing the marker on a surface in order to grab a pointing device. The result depends on the algorithm that recognizes the shaking gesture.

Recognition of a Shaking Gesture: When implementing a shaking gesture, we first have to analyze, how such a movement of the controlled object (i.e., the marker) can be described. We use a statistically-based approach. Fig. 4 shows two trajectories of the origin P_{Ti} of a tracked object over the last half second. The left trajectory (blue lines) comes from a typical non-critical hand movement. The right trajectory shows a shaking gesture. We take two attributes of each trajectory to recognize shaking. The first attribute is the length of the trajectory l_T for the last half second, calculated by summing up the distances between the sampling points P_{Ti} of the trajectory. The second attribute is the spatial spread within this trajectory s_T . We calculate the spatial spread by computing the arithmetic mean of the distances from the sampling points P_{Ti} to their centroid P_c , i.e., the variance.

Comparing those two attributes in a scatter plot (Fig. 5) where the x-axis is the length of the trajectory l_T and the y-axis is the spatial spread within the trajectory s_T , we see that for a shaking gesture, the plotted point goes to the lower right of the scatter plot. This is because, due to the relatively large movement of the tracked object, the length of the trajectory is large, but since all points P_{Ti} are close to each other, the spatial spread s_T is low. So we have defined an area of acceptance which lies below the blue and green line in the scatter plot. Whenever the currently calculated values of the trajectory length l_T and the spatial spread s_T generate a point in that area, a shaking gesture is in progress. To avoid continuous triggering of shaking gestures while an object is still being shaken, we have introduced a dynamic hysteresis for the spatial spread value.

Hystereses used hitherto apply fixed boundaries. With such hystereses it could happen that multiple shaking gestures, rapidly made one after each other, were recognized as only one shaking gesture,

because the spatial spread does not rise above the upper boundary. Another problem with a fixed hysteresis is, that gently made shaking gestures are not recognized, because the spatial spread does never fall below the hysteresis boundaries.

We incorporated a dynamic hysteresis by updating these boundaries according to the current value. Thus, whenever the value falls below the lower boundary, the entire interval is lowered such that the lower boundary is at the current value. When the spatial spread value falls below this boundary for the first time, the shaking gesture is triggered and a variable is set telling the system not to trigger the gesture again until it has been unset. When the value exceeds the upper boundary, the entire interval is raised accordingly, and the variable is unset. When the value now falls below the lower boundary again, the next shaking gesture is triggered.

This allows the user to rapidly perform multiple shaking gestures one after each other. Furthermore it also allows the system to recognize heavy shaking gestures and shyly made gestures with the same setup.

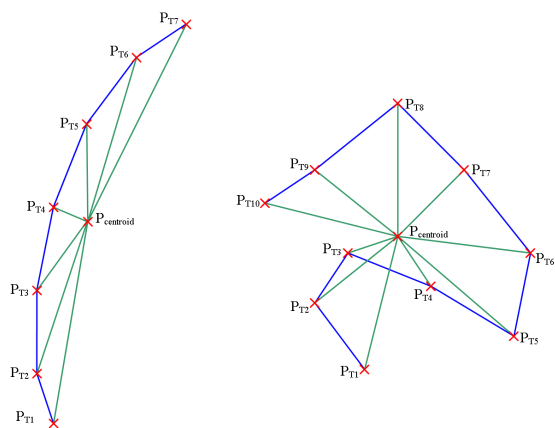


Figure 4: Trajectory length and spatial spread calculation during the past 0.5 second to classify the shaking gesture. left: normal movement, right: shaking

Process to Select a Bond by Toggling Through Two Lists of Atoms: To form a bond, users interact separately with the molecules in their left and right hand. As visual feedback to the user, the currently selected atom of each molecule is shown with pulsating color intensity. To toggle the current selection for either molecule, users briefly shake the respective hand. This can be done in parallel with both hands or sequentially. Whenever they have selected the correct atoms in both molecules, $atom_1 \in A_1$ and $atom_2 \in A_2$, they have specified the tuple $bond = (atom_1, atom_2)$.

3.1.2 Selecting a Bond from a Set of Bonds Between Molecules

As a second method, we use a proximity-based approach to select bonding pairs of molecules. In this case, selection of the two atoms that form tuple $bond = (atom_1, atom_2)$ cannot be treated separately. We measure the pairwise distance between all atoms $atom_1 \in A_1$ with free bonding places of one molecule and the respective atoms $atom_2 \in A_2$ of the other molecule.

To specify a bond, users have to jointly move and rotate the molecules in their left and right hand such that the atoms of the targeted bond form the closest connection. At that point, the respective atoms in both molecules, $atom_1 \in A_1$ and $atom_2 \in A_2$, specify the tuple $bond = (atom_1, atom_2)$, from the set of tuples $B = A_1 \times A_2$.

The shortest bond is shown as a pulsating, semi-transparent cylinder. Optionally, further, longer bonds can also be shown. They

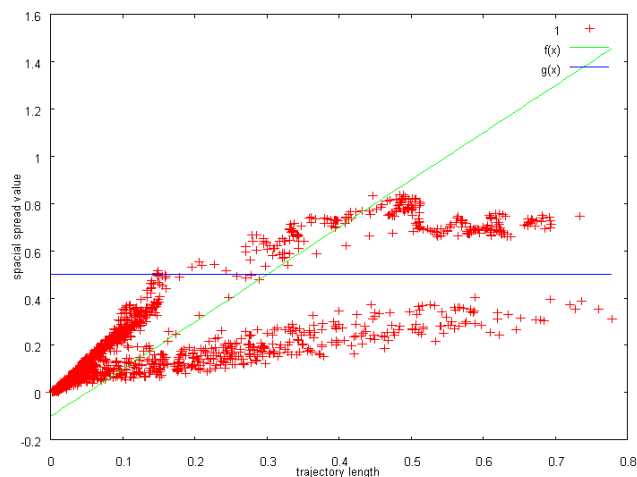


Figure 5: Scatter plot of red points, each showing the current gesture detection value of prerecorded movements. The x-axis shows the length of the trajectory of the last 0.5 second, whereas the y-axis represents the spacial spread value. When a point is below the blue and the green line, a shaking gesture is recognized. The points outside that area are normal movements of the tracked object, thus not triggering a shaking gesture.

are semi-transparent, but they do not pulsate.

As an extension to the proximity-based selection, bonds can also be selected w.r.t. the strength of the attractive force – which is a function of distance as well as other parameters.

3.2 Confirmation of a Selection

Once the user has specified a bond with either the shake-based or the distance-based method, a confirmation signal must be generated to convey to the system that the interactive selection process is now completed.

In order to minimize confounding factors in the current user study, we are using a Wizard-of-Oz technique: users say *Done* when they have selected a bond, and a second person (the experimenter) then hits a button on a keyboard. Other approaches using still gestures exist, but haven't been evaluated and optimized yet.

4 EXPERIMENTAL EVALUATION

The system presented in sections 2 and 3 is operational and has been demonstrated on many occasions. Yet, thus far, it is based on a large number of heuristics and ad-hoc solutions that require further investigation and fine-tuning. We have begun by experimenting with two different methods to specify a selection (as presented in section 3.1.1). We now report on a user study for this issue. All other heuristics were fixed.

4.1 Task

We have conducted a user study to find the differences in speed, precision and user acceptance between the shake-based method and the proximity-based method to select a bond. The participants had to select a series of specified bonds using the shake-based or the proximity-based methods.

4.2 Experimental Setup

We asked 19 people (7 female, 12 male) between the age of 20 and 51 years (mainly between the age of 20 and 30) to participate in the study. The users mainly had no experience with the use of marker based tracking but then stated that they had no problems using it.

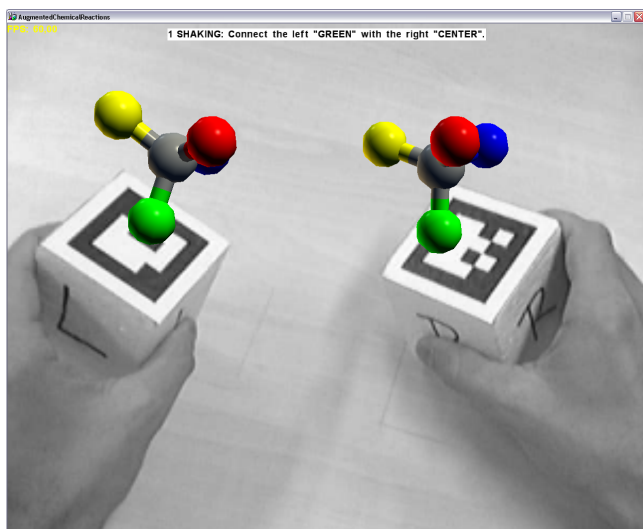


Figure 6: Molecule layout



Figure 7: Experimental setup

The participants interacted with two cube-shaped markers (marked "L" and "R") in the workspace in front of them. Both markers were augmented with identical molecules without chemical semantics. Each molecule consisted of a center atom in gray color, surrounded by 4 atoms in red, green, yellow, and blue (see Figs. 6 and 7).

We used an Intel Core 2 Duo 2.33 GHz Notebook with 2 GB RAM and a NVIDIA GeForce Go 7900 GTX graphics card driven by Windows XP SP3. A Logitech QuickCam Pro 4000 was mounted above the users' head, using a microphone arm. The gray scale camera image with an resolution of 320x240 was displayed and augmented on the display running with a resolution of 1280x1024.

4.3 Test procedure

We first introduced briefly the topic of selecting bonds between two molecules which are directly controlled by the tangible optical markers. Users then had to fill out a first questionnaire, inquiring about age, gender, color blindness and experience with marker tracking.

The participants had to select a specific pair of atoms from two

molecules. The desired combination was shown in an instruction line on the screen in front of their work area, together with the augmented image of the camera above their head, as described in section 2.5.

For the shake-based method, the participants had to select the correct bond by shaking the markers. Each shaking gesture resulted in a change of the selected atom of the respective left or right molecule, cycling through the five atoms in the order of RED, GREEN, YELLOW, BLUE and CENTER. The selected atom was highlighted by letting it pulsate in its transparency. After selecting the two atoms by shaking, participants had to hold the atoms closer than a specified distance to see the bond, shown as a semi-transparent cylinder. We used a Wizard-of-Oz technique for users to confirm their selections: When the user said *Done*, the experimenter confirmed the selection on the keyboard (see 3.1.1). For the performance analysis, the time was taken that the user needed to select a target bond. False combinations were recorded, but not taken into account for the analysis of the needed time.

For the proximity-based method, the molecules had to be moved toward each other to bring up a connection in the form of a semi-transparent cylinder between the nearest atoms of both molecules. When users thought, that the correct combination was established, they had to say *Done* to complete their task. Here also the time was measured, that the participants needed to perform this task.

After each variant, the participants had to fill out a SUS questionnaire [4]. At the end of the study, the users completed a questionnaire to get subjective impressions on each method.

4.4 Test design

We used a within-subject, repeated measures single-session design. Each session lasted about fifteen minutes, including introduction and questionnaires. The session was divided into two parts for both methods. In each part, the participant had to select 24 combinations (5 x 5 atoms, excluding the combination CENTER-CENTER). This set of 24 combinations was shuffled for each user and for each method. After 10 participants the order of the methods was switched to suppress dependencies on a confounding learning effect. Before the first set of 24 combinations users could familiarize themselves with each method by selecting two combinations from each method.

We formulated two hypotheses:

- H1: Proximity based selection will be faster than the shake based approach.
- H2: Shake based selection will be faster than the proximity based approach, when only bonds with the center atom are requested.

4.5 Results

To investigate the performance of both methods, we analyzed both the time participants needed to select the right combination and the error rate of both methods. We also looked at the error rate of our shake recognition algorithm.

4.5.1 Selection Time and Error Rate Analysis

We used a two-tailed t-test for repeated measures on the mean time the users needed to select the correct bond. We found a significant difference in selection time when looking at all performed bonds for $\alpha = 5\%$ ($t(639.08) = 5.358, p < 0.001$), thereby supporting hypothesis H1. When considering all bonds, the proximity based approach was in mean 1.82 seconds faster than the shake based method. In conclusion, hypothesis H2 is accepted. Yet, when looking at only those bonds that involved a (white) center atom, the shake based method was in mean 4.96 seconds faster than the proximity based method ($t(148.85) = 5.804, p < 0.001, \alpha = 5\%$). When analyzing only bonds between outer atoms, on the other hand,

the shake based method was in mean 4.68 seconds slower than the proximity based approach ($t(399.11) = 23.294, p < 0.001, \alpha = 5\%$) (see Fig. 8).

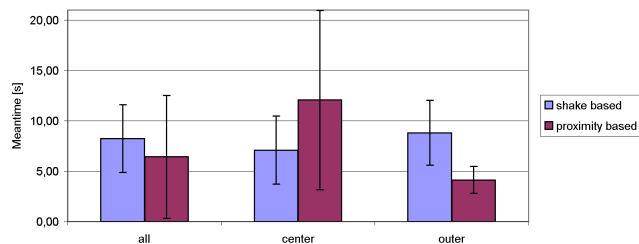


Figure 8: Comparison of the mean selection time between shake based and proximity based approaches with standard deviations

To analyze the error rates, we used a two-tailed t-test on the average errors for each method. An error was made, when a user selected a bond between the wrong atoms. When regarding all bonds, there is a significant difference of 1.58 in mean errors per user for $\alpha = 5\%$ ($t(21.75) = 3.067, p < 0.01$), with the shake based method being worse. Both methods also have a significant difference of 1.47 in mean errors per user for $\alpha = 5\%$ ($t(19.33) = 3.236, p < 0.01$) when considering only bonds with center atoms. Only for bonds using only the outer atoms, there was no significant difference of the average errors per user for $\alpha = 5\%$ (see Fig. 9).

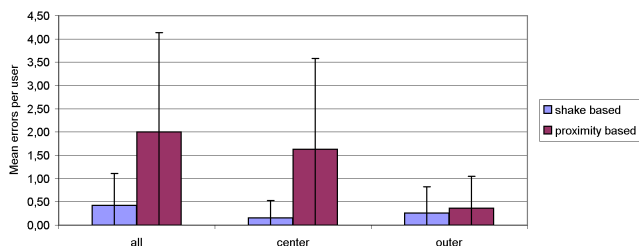


Figure 9: Comparison of the average error rates per user between shake based and proximity based approaches with standard deviations

4.5.2 Error rate analysis of the shake recognition algorithm

To analyze the robustness of our shake recognition algorithm, we recorded all the position and rotation values of all sessions including the timestamps. With this information we were able to replay the movements of the users and analyze them in detail. To calculate the error rate, we counted all shake gestures performed by the users and compared them with the recognized gestures. Unrecognized gestures as well as doubly recognized gestures were counted separately. From the whole of 2265 performed shake gestures, 58 shakes were counted double and 151 were not recognized. This is an overall error rate of 9.2%. In mean, a user performed 119.2 shake gestures, where in mean 2.39% (standard deviation of 2.28%) were recognized double and in mean 6.33% (standard deviation of 6.28%) were not recognized, as seen in figure 10.

4.5.3 Subjective Results

From oral interviews with the test subjects, our thesis was confirmed, that, with the proximity based method, it was very difficult to form a bond of an outer atom with a center atom. The participants also mentioned, that the shaking method was a bit slower but they could select the desired combinations more precisely. In the questionnaire that we gave to the users after the whole session, we

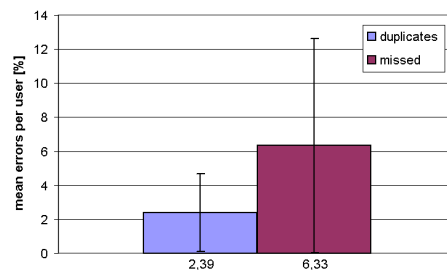


Figure 10: Mean errors in % per user in the shake recognition algorithm of double recognized and not recognized shake gestures with standard deviations.

asked them to give grades for each method on a 6-point Likert scale (with 1=best to 6=worst) for like/dislike, ease of use/difficulty, fast selection/slow selection, accurate selection/inaccurate selection and the difficulty to select combinations with a center atom.

The analysis of the questionnaire shows in Fig. 11 that the users liked both methods and thought that both methods were easy to use, although they thought the shake based approach was a bit easier to use. Users stated that they felt that they were on average equally fast with both methods. Regarding the accuracy and the difficulty to select a center atom for the bonds, the questionnaire reflects that the shake based approach is more accurate and easier to select the center atom.

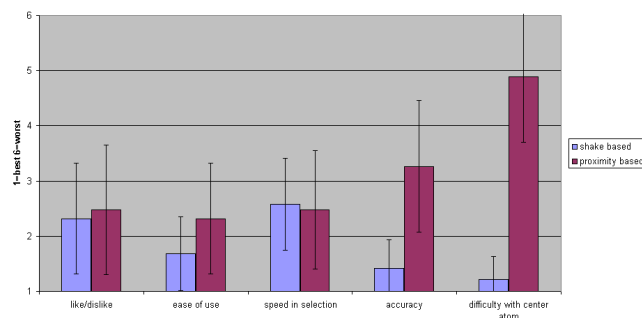


Figure 11: Evaluation of the questionnaire regarding the shake based and proximity based approaches with standard deviations

Having a look at the relatively high SUS-values (Fig. 12), we see that both systems were accepted by the users.

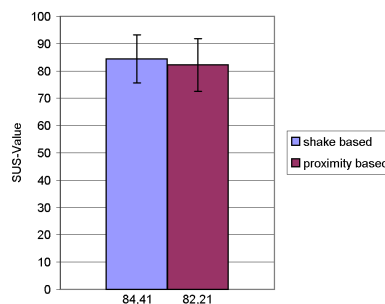


Figure 12: Mean SUS value for the shake based and proximity based approaches with standard deviations

5 DISCUSSION

Combining all results from the experiment, we realize that both implementations have their pros and cons.

When bonding molecules only by atoms on the outer shell, the proximity based approach works very well, since the bonding atoms are easy to reach. When trying to bind with atoms, that are sheltered by surrounding atoms, the proximity based method deteriorates dramatically since the outer atoms often have a shorter distance to the atoms of the other molecule than center atoms and it thus is very hard to establish a bond with center atoms.

Repetitive shaking, on the other hand was considered by some test persons to be increasingly strenuous on the wrists. Furthermore, its performance deteriorates rapidly with increasing numbers of atoms.

It may be possible to combine the advantages of both methods: the preciseness of the shake based method to toggle between very specific, hard to reach options options, and the 3D-immersiveness of the proximity based method to use spatially consistent hand motions to define bonds between the closest atoms. It might be a very good solution to use the shake based method to switch between task contexts on a higher level, i.e. to switch between the inner and outer part of the molecules (or more shells if those exist, or between other kinds of substructures), and then to use the proximity based approach for detailed, direct selection between atoms in the selected sub set.

6 SUMMARY AND CONCLUSIONS

In summary, we have presented a system that uses augmented reality to support chemists in understanding and exploring the dynamic interaction between atoms of several molecules when they are close to one another. AR-based tangible interaction seems to have great potential to combine complex chemical simulations with the opportunities for direct 3D manipulations. Yet, limits may be reached when the specification and selection of new bonds requires users to interact with more than two hands. This limitation might render the entire approach much less usable.

We have addressed this problem by investigating two methods that let users specify new bonds as additional gestures on the side while remaining in the three-dimensional context of directly manipulating the poses of molecules. Both methods have problems, but tests have shown that the methods complement each other and may be combinable to a joint, much better approach.

We conclude that the work has only begun. The first evaluation is encouraging. Yet, many more issues, design aspects and parameter tunings have to be conducted for the system to become a real workhorse for chemists, similar to the many WIMP-based systems that already exist. This is not merely an issue of productization. Rather, it impinges on deep issues of 3D user interaction, such as seamlessly mixing issues of system control (toggling through lists of options) with 3D direct object manipulation. This is the topic of already ongoing and future extensions and evaluations of the system.

ACKNOWLEDGMENTS

We would like to thank all members of FAR for their technical support and for their continuous willingness to see and discuss our demo with us. We would also like to thank the reviewers for their very thoughtful and inspiring comments and suggestions. The project is partially supported by the International Graduate School of Science and Engineering (IGSSE) of the Excellence Initiative of the German federal and state governments (DFG) (www.igsse.tum.de).

REFERENCES

- [1] R. Balakrishnan, G. Fitzmaurice, G. Kurtenbach, and W. Buxton. Digital tape drawing. In *Proc. of Symposium on User Interface Software and Technology*, pages 161–169. ACM New York, NY, USA, 1999.
- [2] R. Balakrishnan and K. Hinckley. Symmetric bimanual interaction. In *Proc. of SIGCHI*, pages 33–40. ACM New York, NY, USA, 2000.
- [3] J. Bartlett. Rock 'n' Scroll Is Here to Stay. *IEEE Computer Graphics and Applications*, pages 40–45, May/June 2000.
- [4] J. Brooke. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, pages 189–194, 1996.
- [5] D. Casalta, Y. Guiard, and M. Lafon. Evaluating two-handed input techniques: Rectangle editing and navigation. In *Proc. of SIGCHI*, page 237. ACM, 1999.
- [6] M. Fjeld, J. Fredriksson, M. Ejdestig, F. Duca, K. Bötschi, B. Voegtli, and P. Juchli. Tangible user interface for chemistry education: comparative evaluation and re-design. In *Proceedings of SIGCHI*, page 808. ACM, 2007.
- [7] M. Fjeld, P. Juchli, and B. Voegtli. Chemistry education: A tangible interaction approach. In *Proc. of International Conference on Human-Computer Interaction (INTERACT)*, page 287. Ios Pr Inc, 2003.
- [8] Gaussian, Inc. GaussView, <http://www.gaussian.com>. accessed Nov. 13, 2009.
- [9] A. Gillet, D. Goodsell, M. Sanner, D. Stoffler, and A. Olson. Augmented reality with tangible auto-fabricated models for molecular biology applications. In *IEEE Visualization*, pages 235–241. IEEE Computer Society, 2004.
- [10] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior*, 19:486–517, 1987.
- [11] S. Güven, S. Feiner, and O. Oda. Mobile Augmented Reality Interaction Techniques for Authoring Situated Media On-Site. In *Proc. of ISMAR*, pages 235–236, Santa Barbara, CA, USA, 2006. IEEE.
- [12] K. Hinckley, R. Pausch, D. Proffitt, and N. Kassel. Two-Handed Virtual Manipulation. *ACM Transactions on Computer-Human Interaction*, 5(3):260–302, 1998.
- [13] HYPERCUBE, Inc. HyperChem, <http://www.hyper.com>. accessed Nov. 13, 2009.
- [14] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proc. IEEE International Workshop on Augmented Reality (IWAR'99)*, pages 85–94, San Francisco, 1999. IEEE.
- [15] J. Looser. *AR Magic Lenses: Addressing the Challenge of Focus and Context in Augmented Reality*. PhD thesis, University of Canterbury. Computer Science and Software Engineering, 2007.
- [16] P. Maier, M. Tönnis, and G. Klinker. Augmented chemical reactions. In *International Conference on Chemical Engineering (ICCE)*, 2009.
- [17] Minnesota Supercomputer Center. Jmol, <http://jmol.sourceforge.net>. accessed Nov. 13, 2009.
- [18] J. Pierce, B. Stearns, and R. Pausch. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proc. of Symposium on Interactive 3D Graphics*, pages 141–145. ACM, 1999.
- [19] R. Paine, Chemistry Dept., Univ. of New Mexico. TINKER Software Tools for Molecular Design, <http://dasher.wustl.edu/tinker/>. accessed Nov. 13, 2009.
- [20] M. Ricknäs. Gestures Set to Shake up Mobile User Interfaces. *PC World*, www.pcworld.com/article/171535, Sep 7, 3:40 pm, 2009.
- [21] Z. Szalavári and M. Gervautz. Using the personal interaction panel for 3d interaction. In *Proc. of Conference on Latest Results in Information Technology*, page 36, 1997.
- [22] S. White, D. Feng, and S. Feiner. Interaction and presentation techniques for shake menus in tangible augmented reality. In *Proc. of ISMAR*, pages 39–48, Washington, DC, USA, 2009. IEEE.
- [23] S. White, L. Lister, and S. Feiner. Visual hints for tangible gestures in augmented reality. In *Proc. of ISMAR*, pages 1–4, Washington, DC, USA, 2007. IEEE Computer Society.
- [24] X. Xia, P. Irani, and J. Wang. Evaluation of Guiard's Theory of Bimanual Control for Navigation and Selection. *Lecture Notes in Computer Science*, 4566:368, 2007.