An Evaluation of an Augmented Reality Multimodal Interface Using Speech and Paddle Gestures

Sylvia Irawati [1, 3], Scott Green [2, 4], Mark Billinghurst [2],
Andreas Duenser [2], Heedong Ko [1]

[1] Imaging Media Research Center, Korea Institute of Science and Technology
[2] Human Interface Technology Laboratory New Zealand, University of Canterbury
[3] Department of HCI and Robotics, University of Science and Technology
[4] Department of Mechanical Engineering, University of Canterbury
{sylvi, ko}@imrc.kist.re.kr
{scott.green, mark.billinghurst, andreas.duenser}@hitlabnz.org

Abstract. This paper discusses an evaluation of an augmented reality (AR) multimodal interface that uses combined speech and paddle gestures for interaction with virtual objects in the real world. We briefly describe our AR multimodal interface architecture and multimodal fusion strategies that are based on the combination of time-based and domain semantics. Then, we present the results from a user study comparing using multimodal input to using gesture input alone. The results show that a combination of speech and paddle gestures improves the efficiency of user interaction. Finally, we describe some design recommendations for developing other multimodal AR interfaces.

Keywords: multimodal interaction, paddles gestures, augmented-reality, speech input, gesture input, evaluation.

1 Introduction

Augmented Reality (AR) is an interface technology that allows users to see three-dimensional computer graphics appear to be fixed in space or attached to objects in the real world. AR techniques have been shown to be useful in many application areas such as education [1], entertainment [2] and engineering [3]. In addition to viewing virtual content, a wide variety of interaction methods have been explored by researchers including using mouse input [4], magnetic tracking [5], real objects [6], pen and tablet [7] and even natural gesture input with computer vision [8]. However, further research on finding the best way to interact with AR content still needs to be conducted, and especially usability studies evaluating the interaction techniques.

This paper presents the design and evaluation of an AR multimodal interface that uses speech and paddle gestures for interaction with virtual objects in the real world. The primary goal of our work is to evaluate the effectivenes of multimodal interaction in an AR environment. This work contributes to the collective knowledge of AR interaction methods by providing an example of a combination of speech and paddle gestures to interact with AR content. It also provides results from a rigorous user study that could be used as guidelines for developing other multimodal AR interfaces.

In this paper, we first review related work and then briefly describe our multimodal architecture. We then discuss our user study and present the results from this study. Finally, we provide design guidelines for the development of multimodal AR interfaces and directions for future research.

2 Related Work

Our work is motivated by earlier research on multimodal interfaces, virtual reality (VR) and AR interfaces. From this research we can learn important lessons that can inform the design of our system.

One of the first multimodal interfaces to combine speech and gesture recognition was the Media Room [9]. Designed by Richard Bolt, the Media Room allowed the user to interact with the computer through voice, gesture and gaze. The user sat in the center of a room with position sensing devices worn on the wrist to measure pointing gestures and glasses with infra-red eye tracking for gaze detection. Computer graphics were projected on the wall of the room and the user could issue speech and gesture commands to interact with the graphics.

Since Bolt's work, there have been many two-dimensional desktop interfaces that show the value of combining speech and gesture input. For example, Boeing's "Talk and Draw" [10] application allowed users to draw with a mouse and use speech input to change interface modes. Similarly Cohen's QuickSet [11] combined speech and pen input for drawing on maps in command and control applications.

Multimodal interfaces can be very intuitive because the strengths of voice input compliment the limitations of gesture interaction and vice versa. Cohen [12, 13] shows how speech interaction is ideally suited for descriptive techniques, while gestural interaction is ideal for direct manipulation of objects. When used together, combined speech and gesture input can create an interface more powerful that either modality alone. Unlike gesture or mouse input, voice is not tied to a spatial metaphor [14], and so can be used to interact with objects regardless of whether they can be seen or not. However, care must be taken to map the appropriate modality to the application input parameters. For example, Kay [15] constructed a speech driven interface for a drawing program in which even simple cursor movements required a time consuming combination of movements in response to vocal commands.

Multimodal interaction has also been used in VR and 3D graphic environments. Early work by Hauptmann [16] employed a Wizard of Oz paradigm and investigated the use of multimodal interaction for a simple 3D cube manipulation task. The study had three conditions; subjects used gestures only, speech only, and gestures and/or speech as they wished. The analysis showed that people strongly preferred using combined gesture and speech for the graphics manipulation.

The QuickSet architecture [11] was integrated into the Naval Research Laboratory's Dragon 3D VR system [17] to create a multimodal system that employs a 3D gesture device to create digital content in a 3D topographical scene. Speech and gesture are recognized in parallel, parsed, and then fused via the Quickset multimodal integration agent. This allowed users to create and position entities by speaking while gesturing in 3D space. Laviola [18] investigated the use of whole-hand and speech input in virtual environments in the interior design. The application allowed a user to create virtual objects using speech commands while object manipulation was achieved using hand gestures. Ciger et al. [19] presented a multimodal user interface that combined a magic wand with spell casting. The user could navigate in the virtual environment, grab and manipulate objects using a combination of speech and the magic wand.

More recent works enhanced the virtual environment by adding semantic models. For example, Latoschik [20] presented a framework for modeling multimodal interactions, which enriched the virtual scene with linguistic and functional knowledge about the objects to allow the interpretation of complex multimodal utterances. Holzapfel et al. [21] presented a multimodal fusion for natural interaction with a humanoid robot. Their multimodal fusion is based on an information-based approach by comparing object types defined in the ontology.

Although AR interfaces are closely related to immersive VR environments, there are relatively few examples of AR applications that use multimodal input. McGee and Cohen [22] created a tangible augmented reality environment that digitally enhanced the existing paper-based command and control capability in a military command post. Heidemann et al. [23] presented an AR system designed for online acquisition of visual knowledge and retrieval of memorized objects. Olwal et al. [24] introduced SenseShapes, which use volumetric regions of interest that can be attached to the user, providing valuable information about the user interaction with the AR system. Kaiser et al. [25] extended Olwal's SenseShapes work by focusing on mutual disambiguation between input channels (speech and gesture) to improve interpretation robustness.

Our research is different from these AR interfaces in several important ways. We use domain semantics and user input timestamps to support multimodal fusion. Our AR system allows the use of a combination of speech, including deictic references and spatial predicates, and a real paddle to interact with AR content. Most importantly, we present results from a user study evaluating our multimodal AR interface.

3 Multimodal Augmented Reality Interface

The goal of our application is to allow people to effectively arrange AR content using a natural mixture of speech and gesture input. The system is a modified version of the VOMAR application [26] based on the ARToolKit library [27]. Ariadne [28] which uses the Microsoft Speech API 5.1, is utilized as the spoken dialogue system.

In the VOMAR application the paddle is the only interaction device. The paddle, which is a real object with an attached fiducial marker, allows the user to make gestures to interact with the virtual objects. A range of static and dynamic gestures is recognized by tracking the motion of the paddle (Table 1).

Our multimodal application involves the manipulation of virtual furniture in a virtual room. When users look at different menu pages through a video see through head mounted display with a camera attached to it (Figure 1), they see different types of virtual furniture on the pages, such as a set of chairs or tables (Figure 2). Looking at the workspace, users see a virtual room (Figure 3). The user can pick objects from the menu pages and place them in the workspace using paddle and speech commands.

Table 1. The VOMAR Paddle Gestures

Static Gestures	Paddle proximity to object Paddle tilt/inclination
Dynamic Gestures	Shaking: side to side motion of paddle Hitting: up and down motion of paddle Pushing object

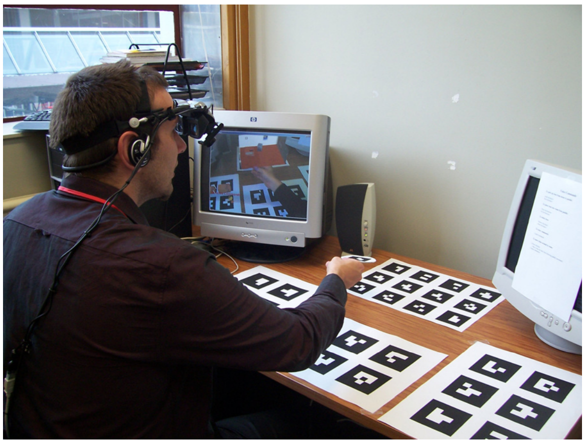

Fig. 1. A participant using the AR system

The following are some speech commands recognized by the system:

- Create Command "Make a blue chair": to create a virtual object and place it on the paddle.
- Duplicate Command "Copy this": to duplicate a virtual object and place it on the paddle.
- Grab Command "Grab table": to select a virtual object and place it on the paddle.
- Place Command "Place here": to place the attached object in the workspace.
- Move Command "Move the couch": to attach a virtual object in the workspace to the paddle so that it follows the paddle movement.

The system provides visual and audio feedback to the user. It shows the speech interpretation result on the screen and provides audio feedback after the speech and paddle gesture command, so the user may immediately identify if there was an incorrect result from the speech or gesture recognition system. To improve user interactivity, the system also provides visual feedback by showing the object bounding box when the paddle touches an object.

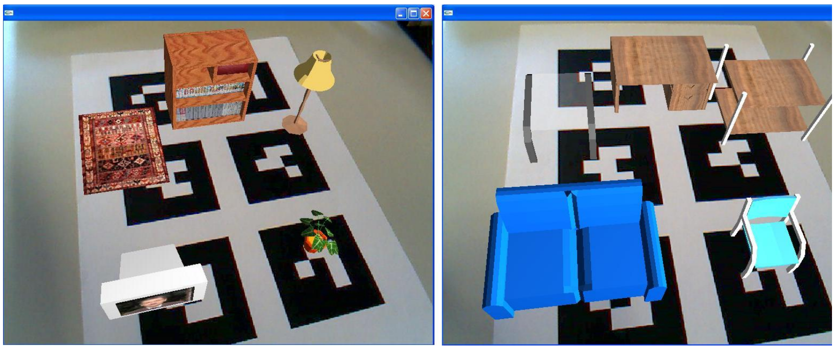

Fig. 2. Virtual menus that contain a set of virtual furniture

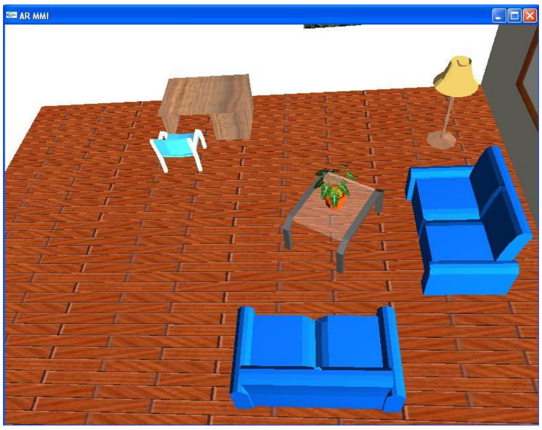

Fig. 3. A virtual room with furniture inside

To understand the combined speech and gesture, the system must fuse inputs from both input streams into a single understandable command. Our multimodal fusion works as follows: when a speech interpretation result is received from Ariadne, the AR Application checks whether the paddle is in view. Next, depending on the speech command type and the paddle pose, a specific action is taken by the system. For example, consider the case when the user says "grab this" while the paddle is placed over the menu page to grab a virtual object. The system will test the paddle proximity to the virtual objects. If the paddle is close enough to an object, the object will be selected and attached to the paddle. If the paddle is not close enough, the object will not be selected.

When fusing the multimodal input, our system also considers object properties, such as whether the object can have things placed on it (defined as `ISPLACEABLE)` or if there is space under the object (defined as `SPACEUNDERNEATH)`. These properties are used to resolve deictic references in the speech commands from the user. For example, if the user says "put here" while touching a virtual couch with the paddle, the possible locations referred to by 'here' are 'on the couch' or 'under the couch'. By checking the object properties of the couch, e.g. `SPACEUNDERNEATH` being false and ISPLACEABLE true, the system understands that 'here' refers to the position 'on top of the couch'. In case the object properties cannot disambiguate user input, the position of the paddle is used by the system. For example, the system checks the paddle in the z (up-down) direction. If the z position of the paddle is less than a threshold value (for example the height of the desk), the system understands 'here' as 'under the desk'.

4 User Study

To evaluate our multimodal AR interface, we conducted a user study. The goal was to compare user interaction with the multimodal interface to interaction with a single input mode. Results from this experiment will help identify areas where the interface can be improved and inform future designs of multimodal AR interfaces.

There were 14 participants (3 female and 11 male) recruited from the staff and students of the HIT Lab NZ. A breakdown of the participants is given in Table 2. The non-native English speakers were foreign-born students who were comfortable speaking English. All male participants used the same male speech recognition profile and all female participants a single female profile. The users did not have to train their own speech recognition profiles. The default profiles proved to be accurate.

Table 2. User breakdown

Criteria	Yes	No
English native speaker	3	11
Familiar with AR	11	3
Familiar with paddle interaction	8	6
Familiar with speech recognition	5	9

Users were seated at a desktop PC and wore a noise canceling microphone and an e-Magin head mounted display with a Logitech Quickcam USB camera attached. The e-Magin display is a bioccular display running at 800x600 pixel resolution with a 26-degree field of view. The Quickcam was capturing 640x480 resolution video images of the real world that were shown in the head mounted display with virtual graphics overlaid onto this real world view. The application was running at 30 frames per second and is shown in Figures 1, 2 and 3 in Section 3.

4.1 Setup

The user study took about forty-five minutes for each user. In the evaluation phase users had to build three different furniture configurations using three different interface conditions;

(A) Paddle gestures only
(B) Speech with static paddle position
(C) Speech with paddle gestures.

To minimize order effects, presentation sequences of the three interface conditions and three furniture configurations were systematically varied between users. Before each trial, a brief introduction and demonstration was given so that the users were comfortable with the interface. For each interface condition the subjects completed training by performing object manipulation tasks until they were proficient enough with the system to be able to assemble a sample scene in less than five minutes. A list of speech commands was provided on a piece of paper, so the user could refer to them throughout the experiment.

Before the user started working on the task a virtual model of the final goal was shown and then the furniture was removed from the scene, with only the bounding box frames remaining as guidance for the user (see Figure 4). The user was also given a color printed picture of the final scene to use as a reference. After performing each task, users were asked questions about the interface usability, efficiency and intuitiveness. After completing all three tasks we asked the users to rate the three interaction modes and to provide comments on their experience. Task completion times and object placement accuracy were recorded and served as performance measures.

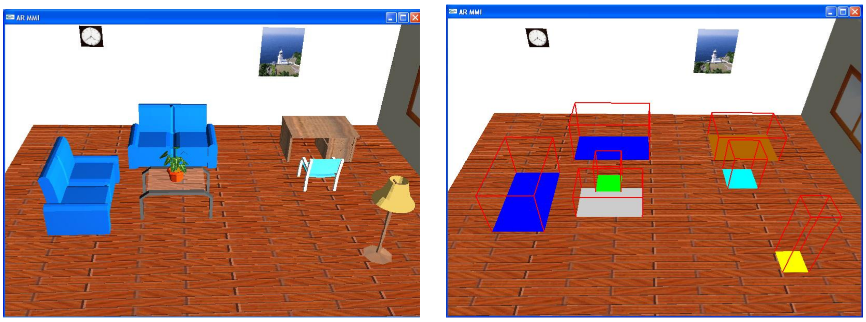

Fig. 4. Final goal and task guidance for the user

4.2 Results and Discussion

Results for average completion times across the three interface conditions are shown in Table 3. Two subjects did not complete the tasks in the time limit of 5 minutes, so they were excluded from the completion time and accuracy analyses.

Table 3. Average performance times

	A: Paddle Gestures Only	B: Speech and Static Paddle	C: Speech and Paddle Gestures
Time (Sec)	165	106	147

When using speech and static paddle interaction, participants completed the task significantly faster than when using paddle gestures only, an ANOVA test finding (F(2,22) = 7.254, p = .004). Completion time for the speech with paddle gestures condition did not differ significantly from the other two conditions. The results show that the use of input channels with different modalities leads to an improvement in task completion time. Part of the performance improvement could be due to the extra time required by the system to recognize and respond to paddle gestures. For example, in the paddle gesture only condition, to drop an object the user had to tilt the paddle until the object slid off. In contrast, using speech the user merely had to say “drop that here” and the object was immediately placed in the workspace.

To measure object placement accuracy, we compared the absolute distance and the rotation angle around the vertical z-axis between the target and final object positions; results are shown in Figures 5 and 6. The analyses shows a significant difference for orientation accuracy ($\chi^2 = 6.000$, df = 2, p = .050) but not for position accuracy ($\chi^2 = 2.167$, df = 2, p = .338). Observing the users, we found that users had difficulty translating and rotating objects using paddle gestures alone. This difficulty was because translating and rotating objects often resulted in accidentally moving other objects in the scene.

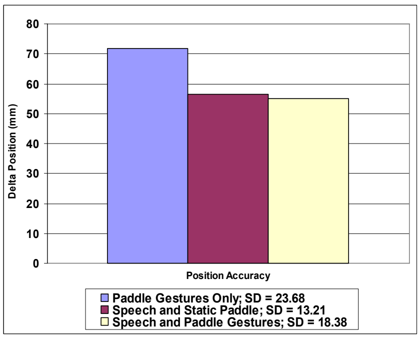

Fig. 5. Result of position accuracy

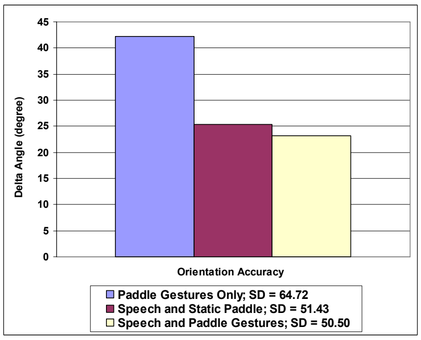

Fig. 6. Result of orientation accuracy

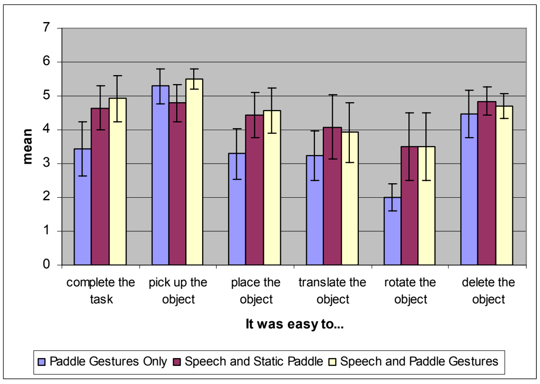

Fig. 7. Result for easiness to do a specific task (95% CI with Bonferroni adjustment)

After each trial, users were given a subjective survey where they were asked on a 6-point Likert scale if they agreed or disagreed with a number of statements (1 = disagree, 6 = agree). Results of this survey are shown in Figure 7. Users felt that completing the task in condition C was easier than condition A ($F(2,26) = 5.554$, $p = .010$). They also thought that placing objects was easier in condition C than in condition A ($F(2,26) = 4.585$, $p = .020$). Users reported that object rotation was easier in conditions B and C than with condition A ($(F(1.152,14.971)_{\text{Huynh-Felt}} = 7.800$, $p = .011$). Thus, users found it hard to place objects in the target positions and rotate them using only paddle gestures. Picking in condition A and C had slightly higher scores than in condition B, although the scores are not significantly different ($(F(1.404,18.249)_{\text{Huynh-Felt}} = 2.893$, $p = .095$). The tendency to prefer picking objects by using paddle gestures shows that users may find this interaction technique quite easy and intuitive.

We also asked the users to rank the conditions from 1 to 3 (with 1 as best rating) in terms of which they liked most (see Table 4). Speech with paddle gestures was ranked highest (mean rank 1.58), then speech with static paddle (mean rank = 1.91), and at last paddle gestures only (mean rank = 2.50). These rankings were significantly different ($\chi^2 = 7.000$, $df = 2$, $p = .030$). This difference could be explained by the observation that users encountered certain difficulties when complementing a specific task. In condition A, most of the users had difficulties in positioning and orienting the objects precisely while in condition B and C the users had better control of the object movement.

Table 4. Mean ranks for conditions

	A: Paddle Gestures Only	B: Speech and Static Paddle	C: Speech and Paddle Gestures
Mean rank	2.50	1.91	1.58

After each experiment was finished we asked the users to provide general comments about the system. Most of the users agreed that it was difficult to place and rotate the virtual objects using only paddle gestures. One user said that pushing the object around using the paddle was quite intuitive but less precise than using the verbal 'move' and 'rotate' commands. Some users suggested adding new gestures to make placement, and especially rotation, of objects easier or to redesign the interaction device (e.g. users should be able to swivel the paddle for easier rotation). Many users were impressed with the robustness of the speech recognition (the system was not trained for individual users) although there were a few users who commented on the difficulties they had in using the speech interface. The users mentioned that accomplishing the task using combined speech and paddle commands was a lot easier once they had learned and practiced the speech commands.

5 Design Recommendations

Based on the observations of people using our multimodal AR interface and the user study results there are some informal design recommendations that may be useful for the design of other multimodal AR interfaces.

Firstly, it's very important to match the speech and gesture input modalities to the appropriate interaction methods. In our case we were using speech to specify commands and gestures to specify parameters (locations and objects) for the commands. It is much easier to say "Put that there" rather than "Put the table at coordinates x = 50, y = 60". The mappings that we used matched the guidelines given by Cohen [13] and others in terms of the strengths and weaknesses of speech and gesture input, and allowed for the use of natural spatial dialogue.

With imprecise recognition based input it is very important to provide feedback to the user about what commands are being sent to the system. In our case we showed the results of the speech recognition on-screen and gave audio feedback after the gesture commands. This enabled the user to immediately recognize when the speech or gesture recognition was producing an error.

It is also important to use a speech and gesture command set that is easy for users to remember. In our case, we only had a limited speech grammar and five paddle gestures. Using combined multimodal input further reduced the amount of commands that users needed to remember; for example it was possible to say "Put that there", rather than "Put the vase on the table".

Finally, the interaction context can be used to disambiguate speech and gesture input. In our case the fusion engine interprets combined speech and gesture input based on the timing of the input events and domain semantics providing two types of contextual cues.

6 Conclusion

In this paper, we describe an augmented reality multimodal interface that uses combined speech and paddle gestures to interact with the system. The system is designed to effectively and easily arrange AR content using a natural mixture of speech and gesture input. We have designed and implemented a test bed application by adding multimodal input to the VOMAR application for the arrangement of virtual furniture. The VOMAR application already had an intuitive tangible AR interface for moving virtual objects using paddle gestures, we enhanced this further by adding speech input.

The results of our user study demonstrate how combining speech and paddle gestures improved performance in arranging virtual objects over using paddle input alone. Using multimodal input, users could orient the objects more precisely in the target position, and finished an assigned task a third faster than using paddle gestures alone. The users also felt that they could complete the task more efficiently. Paddle gestures allowed the users to interact intuitively with the system since they could interact directly with the virtual objects.

Our user study shows that powerful effects can be achieved by combining speech and gesture recognition with simple context recognition. The results also show that combining speech and paddle gestures are preferred over paddle gestures alone. Speech is suitable for control tasks and gestures are suitable for spatial input such as direct interaction with the virtual objects. Contextual knowledge may resolve ambiguous input, in our case, by knowing the object properties, and the position of the paddle, the proper location referred to by the deictic term 'here' can be distinguished.

This is early work and there are several areas of future research that we can explore. The current implementation could be improved by introducing new paddle gestures to optimize the speed, effectiveness and intuitiveness of the interaction, such as gestures for locking/unlocking objects so the user would have more precise control in manipulating the virtual content. The speech grammar could be extended to include more speech commands and dialogue could be added to the system to make the system even more interactive. Finally, this multimodal interface could be extended to other augmented reality application domains to explore if the benefits we have seen in a virtual scene assembly could also be extended to other fields.

Acknowledgments. The first and last authors are funded by the Ubiquitous Autonomic Computing and Network Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea.

References

1. Hannes Kaufmann: Collaborative Augmented Reality in Education. Keynote Speech at Imagina Conference (2003)
2. Istvan Barakonyi, Dieter Schmalstieg: Augmented Reality Agents in the Development Pipeline of Computer Entertainment. In Proceedings of the 4th International Conference on Entertainment Compute (2005)
3. Anthony Webster, Steven Feiner, Blair MacIntyre,William Massie, Theodore Krueger: Augmented reality in architectural construction, inspection and renovation. In Proceedings of .ASCE Third Congress on Computing in Civil Engineering, Anaheim, CA (1996) 913-919
4. Christian Geiger, Leif Oppermann, Christian Reimann: 3D-Registered Interaction-Surfaces in Augmented Reality Space. In Proceedings of 2nd IEEE International Augmented Reality Toolkit Workshop (2003)
5. Kiyoshi Kiyokawa, Haruo Takemura, Naokazu Yokoya: A Collaboration Support Technique by Integrating a Shared Virtual Reality and a Shared Augmented Reality. In Proceedings of IEEE International Conference on Systems Man and Cybernetics (1999) 48-53
6. H. Kato, M. Billinghurst, I. Poupyrev, N. Tetsutani, K. Tachibana: Tangible Augmented Reality for Human Computer Interaction. In Proceedings of Nicograph, Nagoya, Japan (2001)
7. Zsolt Szalavari, Michael Gervautz: The Personal Interaction Panel – A Two-Handed Interface for Augmented Reality. In Proceedings of EUROGRAPHICS, Computer Graphics Forum, Vol. 16, 3 (1997) 335-346.
8. Buchmann, S. Violich, M. Billinghurst, A. Cockburn: FingARtips. Gesture Based Direct Manipulation in Augmented Reality. In Proceedings of 2nd International Conference on Computer Graphics and Interactive Techniques (2004) 212-221
9. Richard A. Bolt: Put-That-There: Voice and Gesture at the Graphics Interface. In Proceedings of the International conference on Computer graphics and interactive techniques, Vol. 14 (1980) 262-270
10. M. W. Salisbury, J. H. Hendrickson, T. L. Lammers, C. Fu, S. A. Moody: Talk and Draw: Bundling Speech and Graphics. IEEE Computer, Vol. 23, issue 8 (1990) 59-65

11. P.R. Cohen, M. Johnston, D.R. McGee, S.L. Oviatt, J.A. Pittman, I. Smith, L. Chen, J. Clow: Quickset: Multimodal Interaction for Distributed Applications. In Proceedings of the Fifth Annual International Multimodal Conference (1997) 31-40
12. P.R. Cohen, M. Dalrymple, F.C.N. Pereira, J.W. Sullivan, R.A. Gargan Jr., J.L. Schlossberg, S.W. Tyler: Synergistic Use of Direct Manipulation and Natural Language. In Proceedings of ACM Conference on Human Factors in Computing Systems (1989) 227-233
13. P.R. Cohen: The Role of Natural Language in a Multimodal Interface. In Proceedings of the fifth symposium on user interface software and technology (1992) 143-149
14. C. Schmandt, M.S. Ackerman, D. Hndus: Augmenting a Window System with Speech Input. IEEE Computer, Vol. 23, issue 8, (1990) 50-56
15. P. Kay: Speech Driven Graphics: a User Interface. Journal of Microcomputer Applications, Vol. 16 (1993) 223-231
16. Alexander. G. Hauptmann: Speech and Gestures for Graphic Image Manipulation. In Proceedings of ACM Conference on Human Factors in Computing Systems (1989) 241-245
17. P. R. Cohen, D. McGee, S. L. Oviatt, L. Wu, J. Clow, R. King, S. Julier, L. Rosenblum: Multimodal interactions for 2D and 3D environments. IEEE Computer Graphics and Applications (1999) 10-13
18. Joseph J. Laviola Jr.: Whole-Hand and Speech Input in Virtual Environments. Master Thesis, Brown University (1996)
19. Jan Ciger, Mario Gutierrez, Frederic Vexo, Daniel Thalmann: The Magic Wand. In Proceedings of the 19th Spring Conference on Computer Graphics (2003) 119-124
20. M.E. Latoschik, M. Schilling: Incorporating VR Databases into AI Knowledge Representations: A Framework for Intelligent Graphics Applications. In Proceedings of the 6th International Conference on Computer Graphics and Imaging (2003)
21. Hartwig Holzapfel, Kai Nickel, Rainer Stiefelhagen: Implementation and Evaluation of a Constraint-based Multimodal Fusion System for Speech and 3D Pointing Gestures. In Proceedings of the 6th International Conference on Multimodal Interfaces (2004) 175-182
22. David R. McGee, Philip R. Cohen: Creating Tangible Interfaces by Augmenting Physical Objects with Multimodal Language. In Proceedings of the 6th International Conference on Intelligent User Interfaces (2001) 113-119
23. Gunther Heidemann, Ingo Bax, Holger Bekel: Multimodal Interaction in an Augmented Reality Scenario. In Proceedings of the 6th International Conference on Multimodal Interfaces (2004) 53-60
24. Alex Olwal, Hrvoje Benko, Steven Feiner: SenseShapes: Using Statistical Geometry for Object Selection in a Multimodal Augmented Reality System. In Proceedings of the second IEEE and ACM International Symposium on Mixed and Augmented Reality (2003) 300–301
25. Ed Kaiser, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Li Xiaoguang, Phil Cohen, Steven Feiner: Mutual Disambiguation of 3D Multimodal Interaction in Augmented and Virtual Reality. In Proceedings of the fifth International Conference on Multimodal Interfaces (2003) 12–19
26. H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, K. Tachibana: Virtual Object Manipulation on a Table-Top AR Environment. In Proceedings of the International Symposium on Augmented Reality (2000) 111-119
27. ARToolKit, http://www.hitl.washington.edu/artoolkit
28. Matthias Denecke: Rapid Prototyping for Spoken Dialogue Systems. In Proceedings of the 19th international conference on Computational Linguistics, Vol. 1 (2002) 1-7