

TeachAR: An Interactive Augmented Reality Tool for Teaching Basic English to Non-Native Children

* Che Samihah Che Dalim^{1,2,3} † Arindam Dey¹ ‡ Thammathip Piumsomboon¹

§ Mark Billingham¹ ¶ Shahrizal Sunar²

¹Empathic Computing Lab
University of South Australia

²MaGIC-X UTM-IRDA
Universiti Teknologi Malaysia

³Faculty of Computer Science
and Information Technology
Universiti Tun Hussein Onn Malaysia

ABSTRACT

Teaching English to children who do not come from an English speaking background is an interesting challenge for educators. In this paper, we present an Augmented reality (AR) tool, TeachAR, for teaching basic English words (colors, shapes, and prepositions) to children for whom English is not a native language. In a pilot study we compare our AR system to a traditional non-AR system. The results indicate a potentially better learning outcome using the TeachAR system than the traditional system. It also showed that children enjoyed using AR-based methods. However, it also showed a few usability issues with the TeachAR interface, which we will improve on in the future.

Keywords: Augmented Reality, Teaching and Learning, English Language, Children, Non-Native Speakers

1 INTRODUCTION

Previous research has shown that Augmented Reality (AR) could be a useful tool for education. Hornecker and Dünser (2007) showed how AR enhanced books could help children with reading difficulties [9], while Kaufmann et al. (2005) showed how AR could be used to teach geometry [12]. Billingham and Dünser provide a high level overview of how AR could enhance traditional learning models [5]. Augmented Reality can broaden childrens learning activities by expanding the boundary of traditional approaches and enhancing visualization of abstract concepts, especially for learning spatial relationships.

In our research we are particularly interested in exploring how AR could be used to teach basic English to young children who are non-English speakers. Previous research has shown how handheld AR could be used to teach Japanese nouns to adults [22], and immersive VR could be used to teach words for spatial relationships in Japanese [20]. However there has been little research on how the technology could be used with young children who have less experience with technology.

The main goal of our work is to explore whether or not Augmented Reality can be used to teach English terms for colors, shapes, and spatial relationships to non-English speaking children. In our work, we have set two main objectives; to compare the effectiveness of using AR to a non-AR application, and to explore

if AR cues combined with speech input enable non-English speaking children to learn English terms for colors, shapes, and spatial relationship more effectively than with the non-speech input.

In the rest of the paper we first provide an overview of earlier related work, then describe the prototype system we have developed, and results from a pilot study with the prototype system. Finally we provide some conclusions and directions for future research.

2 RELATED WORK

Second language learning has been an active area of research for decades. With the use of computers, the methods for language learning have experienced a huge transformation. There has been a transition of learning approaches from traditional teacher-based instructions to modern computer-based learning. Researchers have also begun to study AR and Virtual Reality (VR) interfaces for language learning.

In 1995, Rose and Billingham [20] designed and developed Zengo Sayu, an immersive interactive VR environment to teach Japanese prepositions to non-native adults. The system used voice and gestural commands and user testing found that it was an effective way to teach simple Japanese phrases. Since then, Lin and Lan reviewed 29 articles investigating the use of VR for teaching languages [16]. They reported that between 2004 and 2008 VR systems for language learning were primarily targeted towards higher education groups. However, between 2009 and 2013 a gradual shift of interest was noticed towards junior and senior high school students. They also noted that the interest towards elementary school groups was greatly diminished in the later years. In our study, we are primarily targeting young children 4 to 6 years old and using an AR interface.

One of the first uses of AR for language learning was the work of Wagner [22] who developed a handheld AR application for Japanese Kanji learning. Users could look at Kanji flip cards and see virtual objects representing the nouns shown on the card. However there was no user evaluation conducted with the system, and the software was targeting adult users. Similarly, Juan et al. [10] developed a marker-based AR game for teaching Spanish to children aged 5 to 6 years old. Their study indicated that AR is potentially useful as language learning tools for children. A marker-based AR system for English vocabulary learning was developed by Yuan et al. [7] who conducted a user study finding that students liked the multimedia instruction the AR learning system provided. Tadashi [18] investigated university students brain activities while they learned languages using AR and the traditional printed methods, finding that learning using AR was less stressful than with printed material.

There are some AR applications that aid in real-time language translation and learning. Parhizkar et al. [19] demonstrated an AR mobile translator which can recognize text in Malay language and translate it into English in real time to assist foreigners communi-

*e-mail:samihah@uthm.edu.my

†e-mail:arindam.dey@unisa.edu.au

‡e-mail:Thammathip.Piumsomboon@unisa.edu.au

§e-mail:mark.billinghurst@unisa.edu.au

¶e-mail:shahrizal@utm.my

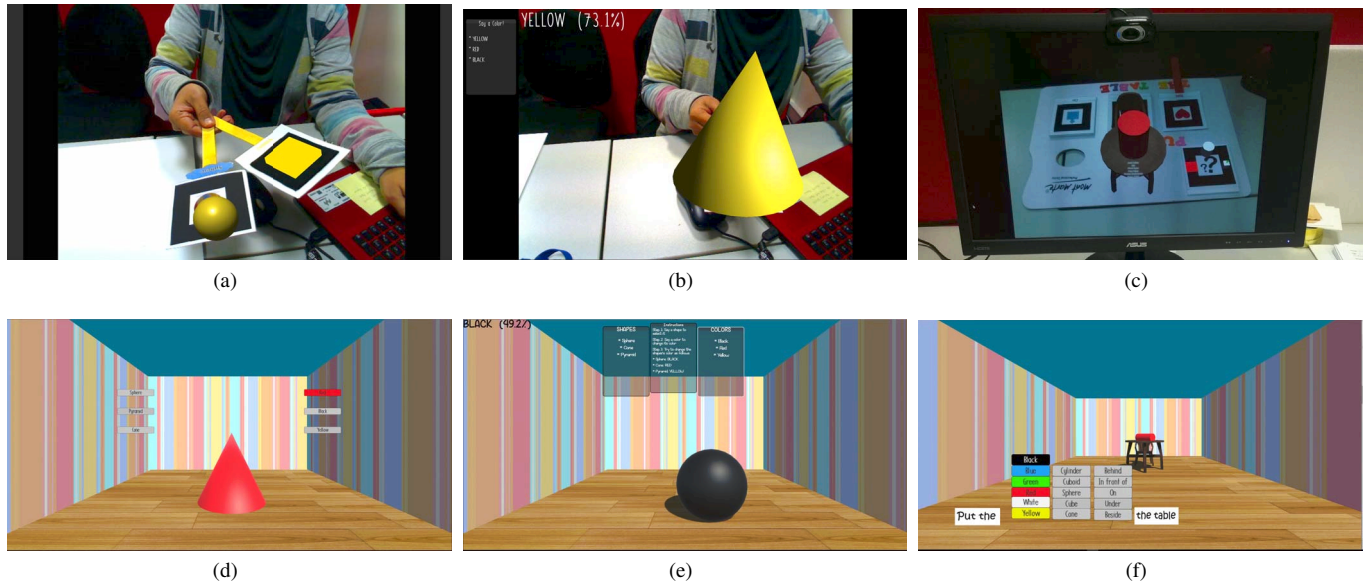


Figure 1: TeachAR (top) vs. non-AR (bottom) modes of learning.

cating with Malay people. Frago et al. [8] also described a fairly similar application called TranslateAR, a multimodal mobile AR translator developed for the Nokia N900 smartphone. A more recent application of this kind was developed by Prema and Madu [17], who developed a system which could detect and translate English words into Telugu in real time. This application aimed to assist students translating English words from their English textbook into Telugu to increase understanding. However, none of the above systems were evaluated with users to identify their acceptance and usefulness.

In recent years, researchers and developers have begun to implement language learning aids in the form of games. Bereira et al. [4] developed Matching Objects and Words, an AR game for learning words in Portuguese and English. This desktop-based game provides visual and auditory cues to motivate elementary school children to memorize how to write and pronounce the names of animals. The target group was children aged 7 to 9 years old. Boonbrahm et al. [6] developed an AR game for English language learning on a mobile platform. It was designed to encourage Thai primary school students to learn written and conversational English. In this game, virtual objects appear in response to written input by the students. Learning conversation was done by watching and listening to virtual characters who speak to each other when two markers were positioned close to each other. However, these systems only provide recorded audio output in order to teach children the pronunciation of words.

Compared to this earlier research our work makes a number of important contributions. Kumar et al. [13] reported that verbal activities such as recalling and vocalizing words for expressing an intended meaning enhances language learning skills. However in the previous AR systems, none consider vocalizing words for language learning. In our AR system, we use speech recognition in the form of voice commands, so that students have the experience of pronouncing words and not only listening to them. This is because we believe that pronunciation is best learnt when children have the experience of saying the word and not only listening to it. In addition, our system is the first AR language learning system to target children below the age of 7 years old. To the best of our knowledge, our work is also the first to use AR for English language learning of 3D shapes and spatial relationships. Teaching abstract concepts such as

spatial relationships is an effective application of AR that should be utilized in education [15, 21]. Shanshan et al. [14] reported based on the result of their pilot study that some participants feel that the words taught in their AR application were concrete enough that the use of AR was not so much helpful. The participants added that it would be more useful if the AR tool is used to teach abstract words which is hardly be imagined through still images. To our knowledge, there has been no work done in teaching spatial relationship for children in AR environment therefore we find it beneficial to teach English words of spatial relationship concepts not only to improve language learning of non-native children, but also to apply AR at its most potential way in education.

3 SYSTEM DESIGN AND ARCHITECTURE

To test the usefulness of AR for English language learning we have developed an educational application called TeachAR and compared it with a non-AR system (Figure 1). TeachAR is a desktop AR application which teaches children about colors, shapes and spatial relationships. Our system uses the Microsoft Kinect [1] for speech recognition, ARToolkit for square marker tracking [11], a 640 x 360 pixel resolution webcam for image capture, and a monitor for viewing the AR scene.

Twelve different markers were used to teach six colors and six shapes (Figure 1a) in the speech-disabled version of TeachAR, with one marker for each color or shape. For the speech-enabled version only one marker was used (Figure 1b) and the virtual object shown on the marker changed its shape and color based on voice commands. To teach prepositions we changed the spatial location of a virtual object in relation to a virtual table shown on the screen (Figure 1c). For the non-AR version (Figure 1d-f), similar changes in the virtual objects were achieved using a mouse click. For example, in Figure 1d the user clicks on the red button on screen to change the color of the virtual cone. We showed buttons on the screen and a list of colors, shapes, and prepositions that can be selected to interact with the system (Figure 1e).

The TeachAR platform can be used in two ways; AR and non-AR. In the AR case children can use AR markers in front of the camera to interactively change the parameters of the virtual objects shown on the marker in the live camera view. In the non-AR case the virtual objects are shown in a 3D virtual scene on the screen and



Figure 2: AR view of Spatial Relationship module.

can be interacted with mouse input.

The TeachAR software comprises of two modules; (1) Color and Shapes, and (2) Spatial Relationships. Each module supports interaction with and without speech input. Interaction by speech means children can speak out loud the name of a color, shape, or the spatial location and the virtual objects will change accordingly. In the interaction without speech case, children can use an AR marker and mouse click to change the parameters of the virtual objects.

To use the software for language learning, children will first use the color and shapes module to learn basic colors and shapes, and then use the second module to learn English terms for spatial relationships.

3.1 Application Development

TeachAR was developed using the Unity game engine [2] with the ARToolkit for Unity plugin. ARToolkit is used for tracking square image markers to create an Augmented Reality view (see Figure 2). A total of 20 square markers were created for the application. In the AR scene rendering is done in Unity which sets the background video to the ARToolkit controller, while the rest of the scene is set as an AR foreground layer in the Unity camera.

The microphone array of the Kinect sensor is used for capturing the users speech and passing it to Microsoft Speech API (SAPI) for recognition based on a list of keywords stored in an XML document. Speech recognition accuracy decreases in a noisy environment and when there are many speakers at the same time. To avoid this problem, we have set the threshold value for the recognition confidence level to 0.4 and only have a single user at a time. The Kinect sensor is positioned 30 cm from children.

4 PILOT EXPERIMENT

We conducted a small pilot study to investigate the effectiveness of the AR-based learning method in comparison to a non-AR learning method. The primary aim of the experiment was to identify usability issues of the system and areas for improvement before running a formal study later with more participants.

4.1 Study Design

We ran a short experiment with four children (two girls, two boys) between the ages of 3 and 6 years old, whose native language are Persian, Korean, Bengali and Malay. Three of the participants have very basic knowledge of English, while one participant only uses his mother tongue to communicate. The eldest participant knew the English term for some of the spatial relationship terms, while



Figure 3: Marker groups.

the rest needed translation into their native language to understand. All of the participants used all experimental conditions (within-subjects). The experiment had two independent variables: Teaching platform (AR vs. non-AR) and Speech Input (On, Off).

The primary purpose was to identify engagement and usability issues in these systems and improve them before running a thorough user study. Accordingly, we primarily focused on subjective feedback and behavioral cues as the primary dependent variable. Additionally, we also had a post-test after each session to find whether the children gained any new knowledge. We asked the children to fill out a short questionnaire before starting the tasks as a pre-test. There we collected information about their knowledge on the colors, shapes, and spatial relationships.

4.2 Procedure and Task

We divided the whole experiment in two modules: (i) Color and Shape learning and (ii) Spatial relationship (prepositions) learning. We taught the following in these modules:

1. **Color:** Red, Green, Blue, Yellow, White, and Black
2. **Shape:** Cube, Cuboid, Cone, Pyramid, Sphere, and Cylinder
3. **Prepositions:** On, Beside, In front of, Behind, and Under

We taught these factors using two different presentation modes: AR and non-AR. In both of these presentation modes children could interact with the system using speech and marker (for AR) or mouse (for non-AR). We did not measure differences in learning between speech and non-speech interactions as it would have required more learning stimuli and more participants to properly counterbalance the experiment.

Participants were first welcomed and asked to fill out a pre-questionnaire where we collected their existing knowledge about colors and shapes. We noticed that all participants knew all the colors that we had in our system. We divided our Color and Shape module in two blocks consisting of three colors and three shapes with no overlap between them (A and B). We counterbalanced the order of AR and non-AR presentation modes. We demonstrated how the application works by explaining what markers are for and how they can be used to change shapes, colors and position of the virtual objects in both ways, using markers and speech. In their first presentation mode, participants learnt the colors and shapes in Block A and in their second presentation mode they learnt spatial relationships in Block B.

In the speech disabled mode, participants show a shape marker to see the virtual shape appear on the marker and by presenting a color marker close to the shape marker, the virtual shape change its default color to the color of the presented marker. In speech enabled mode, participant were given only one marker and they were required to say out loud a shape in order to see it rendered on marker. Once the virtual shape appeared, they may change its color by saying out the color of their preference. After each presentation

mode participants filled out a post-test questionnaire where their newly acquired knowledge was measured.

After the Color and Shape module, participants used the Spatial Relationship module. We asked them to answer a pre-questionnaire to measure their current knowledge about prepositions. In this module, we have grouped the markers into three groups; Colors, Shapes and Preposition groups (see Figure 3), and specified three exact positions for those markers on the playing board so that when the participants place the markers accordingly, it will form a short spatial relationship sentence. For example “Put the *Yellow Sphere* On the table”.

For the speech disabled mode (Figure 3a), participants select a marker from each group and place it correctly on the playing board. The virtual object associated with the markers will first shows up on the preview marker (marker with question mark) in real time. Once all three markers have been detected, our system renders the virtual object on preview marker in the correct position in relation to the table (see Figure 1 (c)). In the speech enabled mode (Figure 3b), only one marker representing each group were used. Based on the marker group, participants have to say out loud a color, shape and spatial relationship in order for the virtual objects to appear on the preview marker. If one object from each of the three groups is detected by our system the virtual object is transferred to the correct table position.

After they used their first presentation mode they answered a post-questionnaire. We did not ask them another post-questionnaire after the second presentation mode as there were not enough prepositions to learn in this pilot study. We video recorded their performance and analyzed them later on to find their level of engagement and happiness (or otherwise).

4.3 Results

Video analysis showed that participants were very engaged and awestruck while using the AR interface. One participant thought that it was a magic trick and he was trying to use his hand gestures to change the shape on the marker. All of them were mostly smiling and wanted to play more while using the AR system. However, sometimes the camera did not properly recognize the marker and the virtual objects were not properly displayed, so the participants felt disengaged. We also noticed that due to the short height of participants they could not always extend their hands far enough to make the marker stay within the field of view of the camera. This indicates a need to design the physical space more carefully and possibly use a camera with a wider field of view.

While using the non-AR mouse-based interface, participants were less engaged and needed more instructions to change between shapes and colors. The force needed to physically press the mouse button was also a concern for one of the participants. All of them were happy when they saw that clicking a button on the screen changed the shape or the color of the object. However, they were less interested in playing longer with the non-AR interface than the AR interface.

In both interfaces, the speech input did not work well. This was because the pronunciations of the participants were not always recognized by the system. Participant speech volume was also relatively low for the system to recognize the speech input. In the future we are planning to place the microphone closer to the participants.

Due to a low number of participants, we did not find a significant difference in new knowledge gain between the two interfaces from the result of the post-test questionnaires. However, we have noticed a slightly higher knowledge gain using the AR interface than non-AR interface. The questionnaires were in the form of matching question (left) to the answer (right). For each correct matching we awarded one point. Out of four participants three improved their score in post-test for Module A (shape and color) with an average gain of 1.25 points (out of a maximum of 6 points). For the Non-AR

interface, three participants improved their score in post-test and one participant deteriorated her score. The average gain was 0.75 point in the non-AR interface. For Module B, one participant using AR interface improved his score by 3 points (out of a maximum of 5 points), with another participant already knew all the prepositions. There was no change in score for participants using the non-AR interface.

We observed that all participants called the activity magic as they saw the added graphics rendered on the marker. One participant said she found it funny when the shapes and colors changed with her voice command and was eager to do more instructions with a louder voice. Children tried to make contact with the virtual objects with every means they could such as using gesture, rotating the marker and lifting the board to see objects that were hidden from view. This shows the effectiveness of the technology in terms of creating a realistic spatial illusion.

At the end of the session, we asked the participants which interface they liked the most. Three participants showed a great interest in the AR Color and Shape module because they can play around with the marker like a magic stick, however, the elder participant said she preferred the spatial relationship module. Although we did not make any interview with the parents, we received a positive feedback from one of the parent as his child begin to use the word “sphere” instead of “circle” to address a sphere shape, and use “under” instead of “down” when referring to an object under a chair at home.

Although these results indicate a higher subjective engagement and improved post-test scores using the AR interface we cannot make a solid inference as only four children participated. With the lessons learned in this pilot test, we will improve the study setup. In the future we will use a between-subject design to avoid fatigue because the participants currently have to undergo many trial conditions, taking a longer time to end. We will also improve our TeachAR interface by testing a different library for AR tracking.

5 CONCLUSION

In this paper we presented TeachAR, an AR system for teaching young children who are non-native English speakers about English terms for basic colors, 3D shapes and spatial relationships. Based on our review of previous studies, our system is the first AR language learning tool attempting to teach young children, 4 to 6 years old, about spatial relationship and shapes.

A pilot study was done to evaluate the usability issue of our system before a more thorough study can be done with more children. The objectives of this study were to see how effective our AR teaching method was compared to a non-AR method, and to explore if the use of speech input was able to further increase the effectiveness of the AR learning system.

Our finding shows that our AR system could be effective as a teaching tool for young children as it enhances the engagement in learning. All participants had no difficulty in interacting with the system even with only one demonstration. The preliminary results show some evidence of learning, and a positive inclination towards using the AR interface over the non-AR interface.

However we also identified some areas for improvements in terms of the study setup and the TeachAR interface design. The AR tracking system needs to be improved to avoid tracking interruption and to reduce confusion for the children. This could be achieved through the use of a non-marker based tracking library like Vuforia [3]. We will conduct a user evaluation to measure differences in learning between speech and non-speech interactions. Finally, we will work on conducting user studies that can explore language learning over time, and how much knowledge retention there is after using the AR system.

REFERENCES

- [1] Kinect for speech recognition. <https://developer.microsoft.com/en-us/windows/kinect>. Accessed: 2016-07-11.
- [2] Unity 3d. <https://unity3d.com/>. Accessed: 2016-07-11.
- [3] Vuuforia. <http://www.vuuforia.com/>. Accessed: 2016-07-11.
- [4] J. Barreira, M. Bessa, L. C. Pereira, T. Ado, E. Peres, and L. Magalhes. Mow: Augmented reality game to learn words in different languages: Case study: Learning english names of animals in elementary school. In *Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on*, pages 1–6, June 2012.
- [5] M. Billinghamurst and A. Dünser. Augmented reality in the classroom. *Computer*, 45(7):56–63, 2012.
- [6] S. Boonbrahm, C. Kaewrat, and P. Boonbrahm. *Using Augmented Reality Technology in Assisting English Learning for Primary School Students*, pages 24–32. Springer International Publishing, Cham, 2015.
- [7] Y.-J. Chang, C.-H. Chen, W.-T. Huang, and W.-S. Huang. Investigating students’ perceived satisfaction, behavioral intention, and effectiveness of english learning using augmented reality. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6, July 2011.
- [8] V. Frago, S. Gauglitz, S. Zamora, J. Kleban, and M. Turk. Translatar: A mobile augmented reality translator. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, pages 497–502, Jan 2011.
- [9] E. Hornecker and A. Dünser. Supporting early literacy with augmented books - experiences with an exploratory study. In *INFORMATIK 2007: Informatik trifft Logistik. Band 1. Beiträge der 37. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 24.-27. September 2007 in Bremen, Deutschland*, pages 555–559, 2007.
- [10] C. M. Juan, E. Llop, F. Abad, and J. Lluch. Learning words using augmented reality. In *2010 10th IEEE International Conference on Advanced Learning Technologies*, pages 422–426, July 2010.
- [11] H. Kato and M. Billinghamurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999. (IWAR ’99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85–94, 1999.
- [12] H. Kaufmann, K. Steinbgl, A. Dnser, and J. Glck. General training of spatial abilities by geometry education in augmented reality. *Annual Review of CyberTherapy and Telemedicine: A Decade of VR*, 2005.
- [13] A. Kumar, P. Reddy, A. Tewari, R. Agrawal, and M. Kam. Improving literacy in developing countries using speech recognition-supported games on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’12*, pages 1149–1158, New York, NY, USA, 2012. ACM.
- [14] C. Y. Li, S. and W. Lafayette. A pilot study exploring augmented reality to increase motivation of chinese college students learning english. *Computers in Education*, 25(1), 2015.
- [15] F. Liarokapis, N. Mourkoussis, M. White, J. Darcy, M. Sifniotis, P. Petridis, A. Basu, and P. F. Lister. Web3d and augmented reality to support engineering education. *World Transactions On Engineering And Technology Education, UICEE*, 3(1):11–14, 2004.
- [16] T.-J. Lin and Y.-J. Lan. Language learning in virtual reality environments: Past, present, and future. *Educational Technology & Society*, 18(4):486–497, 2015.
- [17] P. Meda, M. Kumar, and R. Parupalli. Mobile augmented reality application for telugu language learning. In *MOOC, Innovation and Technology in Education (MITE), 2014 IEEE International Conference on*, pages 183–186, Dec 2014.
- [18] T. Miyosawa, M. Akahane, K. Hara, and K. Shinohara. Applying augmented reality to e-learning for foreign language study and its evaluation. In *Proceeding of the 2012 International Conference on E-learning, E-Business, Enterprise Information Systems, & E-Government*, pages 310–316, July 2012.
- [19] B. Parhizkar, K. Oteng, O. Ndaba, A. H. Lashkari, and Z. M. Gebril. Ubiquitous mobile real time visual translator using augmented reality for bahasa language. *International Journal of Information and Education Technology*, 3(2):124–128, 2013.
- [20] H. Rose and M. Billinghamurst. Zengo sayu: An immersive educational environment for learning japanese. *University of Washington, Human Interface Technology Laboratory, Report No. r-95-4*, 1995.
- [21] B. E. Shelton and R. R. Stevens. Using coordination classes to interpret conceptual change in astronomical thinking. In *Proceedings of the 6th International Conference on Learning Sciences, ICLS ’04*, pages 634–634. International Society of the Learning Sciences, 2004.
- [22] D. Wagner and I. Barakonyi. Augmented reality kanji learning. In *Proceedings of the 2Nd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR ’03*, pages 335–, Washington, DC, USA, 2003. IEEE Computer Society.