cARdLearner: Using Expressive Virtual Agents when Learning Vocabulary in Augmented Reality

Aimée Sousa Calepso*
Natalie Hube*
aimee.sousacalepso@visus.uni-stuttgart.de
natalie.hube@visus.uni-stuttgart.de
University of Stuttgart
Stuttgart, Germany

Noah Berenguel Senn
University of Stuttgart
Stuttgart, Germany
st150100@stud.uni-stuttgart.de

Vincent Brandt
University of Stuttgart
Stuttgart, Germany
st161848@stud.uni-stuttgart.de

Michael Sedlmair
University of Stuttgart
Stuttgart, Germany
michael.sedlmair@uni-stuttgart.de

Figure 1: Screenshot of the Augmented Reality view with the flashcard *tanzen*. The avatar is displaying the animation of the presented verb together with the related emotion.

ABSTRACT

Augmented reality (AR) has a diverse range of applications, including language teaching. When studying a foreign language, one of the biggest challenges learners face is memorizing new vocabulary. While augmented holograms are a promising means of supporting this memorization process, few studies have explored their potential in the language learning context. We demonstrate the possibility of using flashcard along with an expressive holographic agent on vocabulary learning. Users scan a flashcard and play an animation that is connected with an emotion related to the word they are seeing. Our goal is to propose an alternative to the traditional use of flashcards, and also introduce another way of using AR in the association process.

*Both authors contributed equally to this research.

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…$15.00
https://doi.org/10.1145/3491101.3519631

CCS CONCEPTS

• **Applied computing → Interactive learning environments**;
• **Human-centered computing → Mixed / augmented reality**.

KEYWORDS

augmented reality, language learning, contextual learning

ACM Reference Format:
Aimée Sousa Calepso, Natalie Hube, Noah Berenguel Senn, Vincent Brandt, and Michael Sedlmair. 2022. cARdLearner: Using Expressive Virtual Agents when Learning Vocabulary in Augmented Reality. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491101.3519631

1 INTRODUCTION

One of the core difficulties students face when learning a new language is associating vocabulary. Often the mother or base language from which someone is trying to learn does not relate with the target language, which poses a challenge when associating new words. Traditionally, vocabulary learning is often achieved using paper material, like flashcards, or passive multimedia material, as in videos and songs. It is often common to learn foreign languages in a classroom setting, using books, or talking to natives of the target language. Along these lines, one of the factors that can help learning vocabulary is the association of the word with a visual cue. Multimedia Theory [23] investigates this topic more deeply, and states that people learn more from words and pictures than from words alone. However, it is not yet well know how we can make use of augmented spaces with dynamic content in order to support the vocabulary learning process.

When experiencing a situation in life where there is the need of a specific word, it becomes easier to remember it afterwards, because of the connection made between the experience and the meaning of the word [18]. One example would be to order food at a restaurant, where the ingredients describing the dishes are presented in the menu. Immersive applications can facilitate the representation of more dynamic situations, which cannot be achieved with other passive material. Towards enriching this vocabulary learning process, the Augmented Reality (AR) community has started with overlaying words on objects. For instance, some studies that explore how virtual agents can convey different experiences [1, 4, 26, 31] have been performed. Our idea is based on bridging the use of tangible assets, like flashcards, with the potential of AR. A trivial approach would be just projecting content on top of flashcards, like vocabulary translation. However, the digital nature of AR offers a wider range of possibilities.

AR can be used to present the meaning of words without using the written translation alone. In this sense, we try to shed a light on how animated holograms can play a role when learning vocabulary, specially when they are related to actions. One of the assumptions from the Multimedia Learning Theory [23] is that learning only takes place when the individual can connect the information with some prior knowledge. Since emotions are a basic human state, we expect that exploring this factor in the shape of animated virtual agents can foster the association process.

This work contributes with a prototype to integrate AR to learning methods that can be improved with emotional elements. We use AR with paper flashcards, presenting animations situated close to the word. The application serves as a showcase to demonstrate how our concept works. We demonstrate how vocabulary from a foreign language can be linked to a virtual agents' emotion representation, and displayed using holograms in a mobile device. We argue that in addition to improving traditional ways of learning a new language, our proposed method could also bring advantages to neurodivergent people, who often struggles without complementary methods [14, 17, 32]. We offer a reflection based on other literature works as well as some design decisions based on the experience we had developing this prototype. With future user studies, we can investigate more the advantages of the proposed method as well as more concrete design guidelines for applications like this.

2 RELATED WORK

Supportive tools for the language learning process have existed for quite some time, including classic approaches like flash cards. Since the AR field is growing steadily, it opens up more and more opportunities to create even more effective tools. Research shows that AR has the potential to increase the interest, level of commitment and confidence of students [15]. It could also provide opportunities for a more independent learning process [2], improve satisfaction and increase motivation in students [5, 10, 21]. In contrast, there is still more investigation needed in terms of user evaluation. Some research has also been done to find out the potential of AR in terms of inclusion in education[11].

Using a situated vocabulary approach, Santos et al. [29] proposed an AR application to help people learn vocabulary. In their work, they implemented augmented support for learning words in Filipino and German, and evaluated their method through a user study in two experiments, comparing AR with traditional digital flashcards over a few days. They observed that students paid more attention using the AR system. Although still preliminary, their research shows that AR as multimedia may lead to better attention and satisfaction.

Previous work by Draxler et al. [11] used AR to create sentence prompts from circumstantial subjects in the environment. These were then presented in a close quiz about grammar. This approach has the advantage of the ability to dynamically generate learning content anywhere. Recognition of objects in the environment and incorporation into learning material, however, would present significant challenges for our use case—especially given that our application would require the recognition of verbs instead of nouns.

Another input modality that has seen use in augmented reality is that of speech. Control via speech seems like a very fitting input technique for our language learning application, and while the pronunciation of words is an important aspect of learning a language,

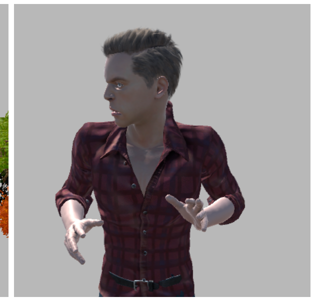
Table 1: List of verbs used in the application together with their translation and related emotion.

English	German	Related emotion
(to) dance	tanzen	happy
(to) cry	weinen	sad
(to) exercise	trainieren	neutral
(to) laugh	lachen	happy
(to) sing	singen	happy
(to) sicken	ekeln	disgust
(to) startle	erschrecken	surprised
(to) stomp	stampfen	angry
(to) panic	durchdrehen	fear

(a) ***(to) dance & happy***

(b) ***(to) cry & sad***

(c) ***(to) exercise & neutral***



(d) ***(to) sicken & disgust***

(e) ***(to) stomp & angry***

(f) ***(to) startle & be surprised***

(g) ***(to) panic & fear***

Figure 2: The screenshots display some example verbs together with their related emotion. We used the six basic emotions by Ekman [13] and added a *neutral* baseline emotion.

we opted not to add this aspect because of technical challenges that arose in previous works [8].

Finally, our solution follows other previous works [8, 25] by using marker cards placed on a surface as an input mechanism for selecting words.

3 METHODS AND APPROACH

To facilitate the implementation of the animations and their relation to emotions, we selected verbs as our target word class. We selected ten German verbs and implemented one animation to each. Our agent is a minimalist human miniature that reacts with body movements depending on the verb selected. We also chose to have printed flashcards that can be recognized by the app through the mobile device camera. Our approach to this problem is to involve the user more in the word formation process and make it more tangible, by creating a direct analogy in the augmented reality world. Table 1 shows the verbs we selected to showcase the concept.

In our implementation, flashcards containing a verb are recognized by the device's camera. We expect that the increased tangibility and immediacy of augmented reality improves memorability of the verb–meaning associations.

The AR approach in itself has the advantage of making the concepts embodied by the verbs intuitively palpable, thus breaking the limits of what traditional learning media such as books or cards can achieve. Specifically, it requires direct interaction, unlike videos, and places the scene directly in front of the user, unlike completely virtual worlds. The intention behind this is to create a more interesting and appealing learning environment compared to more conventional learning methodologies. Here, AR also brings the advantage of situativity, placing the virtual content relative to the position of the cards in the real world.

3.1 Virtual agents and Emotion

We used a miniature human body representation as an agent for our situated visualization of the cards. Thus, the agent's movement and facial expressions need to be explicit, else wise even exaggerated, to sufficiently visualize the verbs otherwise they might be overlooked. Conveying emotions through a virtual agent can be helpful to strengthen meanings transported with actions. Different approaches exist to systematically describe facial expressions that represent emotions while the Facial Action Coding System (FACS) [12] is possibly the best-known. We decided to use the six basic emotions [13] as they are the most distinguishable and selected verbs that link to the displayed emotions of the agent. Compared to the body movement itself, animating the face of the miniature was done using more exaggerated emotional expressions to make sure the emotions can be perceived using a mobile device.

3.2 Prototype Description

The AR application was implemented with Unity [33] and Vuforia [27]. Our miniature was created using the Character Creator 3 that exports with a fully rigged avatar and facial blend shapes. For the body movement, we used animations from Mixamo [19] and retargeted them to fit our miniature rig. Based on FACS [12], we created the emotions shown in Table 1.

The application can be used by mobile devices that support the usage of AR. When using the application, a 3D scene is loading asking the user to use an AR marker. As cards would most likely be placed onto an empty table, we also added background scene decoration (as seen in Figure 3). Our concepts included vast background scenes to provide an appropriate atmosphere. The animated environment was created using Blender[6].

3.3 Menu Navigation

Despite its focus on AR, our application still makes use of a traditional two-dimensional user interface for getting started with the AR scene and for viewing auxiliary information.

The users can first select in a main menu whether they want to learn the German verbs, view the tutorial, or to quit the application. If the user chooses to learn verbs, they will see another selection screen, showing all available verbs grouped into days. The user can then select either the current day to learn new verbs, or select the previous days to refresh their knowledge of already learned verbs, or to experience previous AR scenes again.

If the users selected the tutorial, they will see a new screen where they see multiple images with a corresponding description. Each of them describes one important aspect of our approach and how to use our application. The user can return to this tutorial at any time. This is useful for participants of a future study, in case they forget how the system work. Since we want to minimize the interference with our participants during the individual learning process, we added the tutorial, so users have more freedom, and we do not break their learning immersion with external interference.

3.4 Learning Environment

After a day is selected, the interactive AR learning phase begins. During this phase the users will see the view of the device camera on the screen. The application can support stationary, hand-held, and potentially also head-mounted operation with a few minor adjustments. As we develop the application using Unity, there is the support for a variety of deployment targets, which would make it possible to run it on a webcam-equipped computer, a mobile device, or an augmented-reality HMD (Head-Mounted Display), for instance.

3.4.1 User Interface. In addition to the camera view, a few fixed 2D interactable UI elements are constantly presented to the user. One allows the user to go back to the day selection screen, while the others are there to help the user during the learning process.

Finally, in case the user cannot infer the meaning of the verb from the AR scene, we also display a small translation as a hint in the bottom right corner. This should just be used as additional help to also mentally map the seen contents of the AR scene to something in written form.

3.4.2 Visualization of Verbs. To visualize the verbs, we strove to use the most obvious and simple visual analogy, providing just enough context to be able to disambiguate the meaning. First and foremost, it was important to make the animations for each verb very distinct, such that users would not mistake one for another or mix them up.

3.4.3 Designing Verbs Marker. Our markers were designed with Sketch [30]. For this, we used a pattern from Brovision [7] as a background pattern. The different markers are used as an image target by Vuforia in Unity. The markers can be easily detected using the camera of a mobile device.

Figure 3: Overview of the AR scene. The suggested area for placing the verb cards is indicated by the blue circle.

4 DISCUSSION

For the past years, considering AR applications for education, the more prevalent ones are in the fields of natural sciences and math [16, 28]. Therefore, other fields are yet to be more explored. The process of learning languages is a complex one, which involves not only vocabulary, but grammar, different phonemes and sometimes an entirely different alphabet. AR can be an advantage in some of these factors, like labeling real objects with their foreign names [11]. It is possible to use the situativity aspect to project different types of information close to real objects, which opens up a wide range of possibilities, that are yet to be explored. Applications requiring less technical preparations, like object recognition training or marker placement, would be very welcome in settings like a classroom. In this sense, including the recognition of words, which are already used in flashcards, is a good design choice.

Another advantage to AR applied to language learning is the inclusion of people with disabilities or cognitive impairment. Applications used to support deaf people have also been developed [3, 20, 24], which makes sense due to its visual immersive aspect. Immersive environments have been used for children education and also seems to be a good direction for neurodivergent children [22]. Daniel, et al. [9] implemented application involving emotion recognition by autistic people in immersive environments. In these cases, the design choices have to be careful made taking into consideration the user and their needs.

While developing our app, some design decisions have to be made, as well as some considerations regarding the application scope and our limitations.

4.1 Extensibility

An additional consideration is that as more verbs are added, animations and meanings may overlap and become more confusing for the user. In this case, it may be sensible to add more interactivity, such as the idea of making the user perform parts of the action to further distinguish verbs from one another and also improve memorability. For our initial prototype we only considered the German language, which has this compound verb feature, but it would be interesting to explore different aspects of other languages where AR could take an advantage from. However, for future extensibility, we also would like to consider machine-learning approaches to generate animations for different words in a more sustainable way.

4.2 Word Selection

We acknowledge that the selection of words used to showcase our concept is short. One of the main reasons for this is that creating animations is a time-consuming task, and the more words included, the more animations would have to be developed. Another time-consuming task is to classify each verb into one animation. However, in an immersive learning scenario, the emotions related to the words would not have to be so explicit and could be only a part of the whole environment.

Furthermore, a consideration taken into account in our prototype is that verbs may have different meaning depending on context, and that especially transitive verbs require an object that may benefit from clearer designation as not being part of the meaning itself. We acknowledge that we did not explore other types of vocabulary or verbs that are not strongly related to an emotion, but we also think that the emotion factor is just one of the highlights that can be explored using AR.

4.3 Evaluation and Study Design

Different demographics may cause different results in a future evaluation, so it may be interesting to find correlations especially regarding age, and, for children, whether they are intrinsically or extrinsically motivated using our learning tool. This might be an important factor, because of our central use of AR in an interactive or almost playful manner. We are interested in conducting a future user study, and specially with learning we have to be careful regarding its design. In an ideal scenario we would have users from different ages and language background, and we would also like to conduct it as long term use study, maybe offering the application to be used by students for a longer period of time than a lab study would allow. It would be also interesting to compare the ecological validity of this approach having two different groups, a control one and another using the app, and see how AR would impact the learning process.

4.4 AR role in learning

AR should not be seen just as another screen where we can project information. Our role as researchers is to explore what can go beyond the traditional interaction methods that we already use. The interaction aspect here does not involve the physical body, but it is intrinsic to the cognitive aspect of learning. Going beyond the interaction factor as, for instance, just deciding the user position or the selection technique is part of what we try to explore with this paper. The emotional facet of interaction is harder to evaluate, but we would like to start exploring AR beyond just being the extension of pen and paper or as the augmentation of a screen. In addition, education should not be strictly a top to bottom process, such as one of a classic classroom, and to make feasible other kind of approaches is to open the eyes to a new way of using technology, empowering the users.

5 CONCLUSION AND FUTURE WORK

Since the integration between AR and methods to learn vocabulary has not yet been widely explored, in this work we proposed a prototype to add to this discussion. We propose a different approach to the use of AR with flashcards in the context of vocabulary learning. We showed how a virtual agent could have emotion-related animations that can be connected to vocabulary and increase association capabilities. Demonstrating the potential of the prototype, we hope to pave the way to the integration of AR with another crucial application, that is relevant to when learning a new language. Other features can be incorporated in the future, like gamification elements, and Artificial Intelligence algorithms. Options for future work would involve, for instance, presenting the grammatical gender of nouns of the objects through color coding. Another aspect of this work that demanded a lot of time was the association of verbs with the emotions. In this sense, it could be an alternative to use algorithms to categorize the verbs into emotions, even when not so "strong", which would reduce the issue of manually sorting the verbs. Another use of intelligent algorithms could be recognition of sentences that would trigger responses from the virtual agent. We also would like to further explore the emotion expression including other objects and elements apart from the agent alone.

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618, as well as EXC 2075 - 390740016.

REFERENCES

[1] Ali Abbas, Minji Choi, JoonOh Seo, Seung Hyun Cha, and Heng Li. 2019. Effectiveness of immersive virtual reality-based communication for construction projects. *KSCE Journal of Civil Engineering* 23, 12 (2019), 4972–4983.

[2] Murat Akçayır and Gökçe Akçayır. 2017. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review* 20 (2017), 1–11.

[3] A Almutairi and S Al-Megren. 2017. Preliminary Investigations on AR for the Literacy Development of Deaf Children. *Mayalysia: Springer International Publishing* (2017).

[4] Sahar A. Aseeri and Victoria Interrante. 2018. The Influence of Avatar Representation and Behavior on Communication in Social Immersive Virtual Environments. In *IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*. 823–824. https://doi.org/10.1109/VR.2018.8446261

[5] Jorge Bacca, Silvia Baldiris, Ramon Fabregat, et al. 2018. Insights into the factors influencing student motivation in augmented reality learning experiences in vocational education and training. *Frontiers in psychology* 9 (2018), 1486.

[6] Blender Foundation. 2022. Blender. https://blender.org. Online; accessed 12 January 2022.

[7] Brosvision. 2022. Augmented Reality Marker Generator. https://www.brosvision.com/ar-marker-generator/. Online; accessed 12 January 2022.

[8] Che Samihah Che Dalim, Arindam Dey, Thammathip Piumsomboon, Mark Billinghurst, and Shahrizal Sunar. 2016. TeachAR: An interactive augmented reality tool for teaching basic English to non-native children. In *IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, 82–86.

[9] Vicente Lopez Trompo Daniel, Han Ting, Ratsamee Photchara, and Takemura Haruo. 2019. An AR Puzzle Application for Improving Emotion Recognition for AS Children. In *Proc. of Int. Conf. on Digital Technology in Education*. 52–56.

[10] Ángela Di Serio, María Blanca Ibáñez, and Carlos Delgado Kloos. 2013. Impact of an augmented reality system on students' motivation for a visual art course. *Computers & Education* 68 (2013), 586–596.

[11] Fiona Draxler, Audrey Labrie, Albrecht Schmidt, and Lewis L Chuang. 2020. Augmented reality to enable users in learning case grammar from their real-world interactions. In *Proc. of the Conference on Human Factors in Computing Systems (CHI)*. 1–12.

[12] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior* (1978).

[13] Paul Ekman and Erika L Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

[14] Marc Fabri, Salima Y Awad Elzouki, and David Moore. 2007. Emotionally expressive avatars for chatting, learning and therapeutic intervention. In *Int. Conf. on human-computer interaction*. Springer, 275–285.

[15] Javier Fombona, Maria-Angeles Pascual, and MariCarmen González-Videgaray. 2017. M-learning and Augmented Reality: A Review of the Scientific Literature on the WoS Repository. *Grupo Comunicar* (2017), 63–72.
[16] Juan Garzón, Juan Pavón, and Silvia Baldiris. 2017. Augmented reality applications for education: Five directions for future research. In *Int. Conf. on augmented reality, virtual reality and computer graphics*. Springer, 402–414.
[17] Jonas Grund, Moritz Umfahrer, Lea Buchweitz, James Gay, Arthur Theil, and Oliver Korn. 2020. A Gamified and Adaptive Learning System for Neurodivergent Workers in Electronic Assembling Tasks. In *Proc. of the Conference on Mensch Und Computer* (Magdeburg, Germany) *(MuC '20)*. Association for Computing Machinery, New York, NY, USA, 491–494. https://doi.org/10.1145/3404983.3410420
[18] Evelien Heyselaar, Peter Hagoort, and Katrien Segaert. 2017. In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior research methods* 49, 1 (2017), 46–60.
[19] Adobe Systems Incorporated. 2022. Mixamo Animations - Motion Pack. https://www.mixamo.com/#/?page=1&type=Motion. Online; accessed 12 January 2022.
[20] Chutisant Kerdvibulvech. 2016. A novel integrated system of visual communication and touch technology for people with disabilities. In *Int. Conf. on Computational Science and Its Applications*. Springer, 509–518.
[21] Tsung-Yu Liu and Yu-Ling Chu. 2010. Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education* 55, 2 (2010), 630–643.
[22] Anabela Marto, Henrique A Almeida, and Alexandrino Gonçalves. 2019. Using augmented reality in patients with autism: A systematic review. In *Thematic Conference on Computational Vision and Medical Image Processing (ECCOMAS)*. Springer, 454–463.
[23] Richard E Mayer. 2002. Multimedia learning. In *Psychology of learning and motivation*. Vol. 41. Elsevier, 85–139.
[24] Mohammad Reza Mirzaei, Seyed Ghorshi, and Mohammad Mortazavi. 2014. Audio-visual speech recognition techniques in augmented reality environments. *The Visual Computer* 30, 3 (2014), 245–257.
[25] Tadashi Miyosawa, Mayuko Akahane, Kentaro Hara, and Kikunori Shinohara. 2012. Applying augmented reality to e-learning for foreign language study and its evaluation. In *Proc. of the Int. Conf. on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*. The Steering Committee of The World Congress in Computer Science, 1.
[26] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. *Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173620
[27] PTC, Inc. 2022. Vuforia Engine. https://developer.vuforia.com/. Online; accessed 12 January 2022.
[28] Jairo Quintero, Silvia Baldiris, Rainer Rubira, Jhoni Cerón, and Gloria Velez. 2019. Augmented reality in educational inclusion. A systematic review on the last decade. *Frontiers in psychology* 10 (2019), 1835.
[29] Marc Ericson C Santos, Takafumi Taketomi, Goshiro Yamamoto, Ma Mercedes T Rodrigo, Christian Sandor, Hirokazu Kato, et al. 2016. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning* 11, 1 (2016), 1–23.
[30] Sketch B.V. 2022. The digital design platform - Sketch. https://www.sketch.com/. Online; accessed 12 January 2022.
[31] Harrison Jesse Smith and Michael Neff. 2018. *Communication Behavior in Embodied Virtual Reality*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173863
[32] Lui Thomsen and Ali Adjorlu. 2021. Designing a collaborative virtual reality system to assess social inclusion among neurodiverse students. In *IEEE Conf. on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 353–357. https://doi.org/10.1109/VRW52623.2021.00070
[33] Unity Technologies. 2022. Unity. https://unity.org. Online; accessed 12 January 2022.