



## Gesture Based Word (Re)Acquisition with a Virtual Agent in Augmented Reality: a Preliminary Study

---

Manuela Macedonia, Joachim Greiner, Claudia Repetto and  
Selina Wriessnegger

EasyChair preprints are intended for rapid  
dissemination of research results and are  
integrated with the rest of EasyChair.

October 30, 2021

# Gesture based word (re)acquisition with a virtual agent in augmented reality: A preliminary study

Manuela MACEDONIA<sup>a, 1</sup>, Joachim GREINER<sup>b</sup>, Claudia REPETTO<sup>c</sup>, Selina WRIESSNEGGER<sup>b</sup>

<sup>a</sup> *Dep. of Information Engineering, University, Linz, Austria*

<sup>b</sup> *Dep. of Neural Engineering, Technical University, Graz, Austria*

<sup>c</sup> *Dep. of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy*

**Abstract.** From an evolutionary perspective, language and gesture belong together as a system serving communication on both an abstract and a physical level. In aphasia, when language is impaired, patients make use of gestures. Laboratory research has provided evidence that gesture can support aphasia rehabilitation, more specifically anomia rehabilitation. Here, we test an anomia gesture-based rehabilitation scenario with a virtual trainer (VT) in augmented reality (AR) as a therapy simulation. Thirty German speaking participants were trained to 27 bi- and three-syllabic words of Vimmi, an artificial language. Each Vimmi word was paired to a function word in German. The participants were divided into two Groups of 15 and 15 persons. Group A learned words pairs by observing the gestures performed by the VT and additionally imitating them. Group B learned 27 word-pairs by observing the VT standing still and listening to it. Participants were trained singularly on 3 days alternating one day of training with one day of rest for memory consolidation. Word retention was assessed immediately after each training session by means of free and cued recall tests administered electronically. Group A and Group B did not differ in word retention. When subdividing participants in high and low performers interactions showed that high performers benefitted more of gesture-based training than low performers. The data in this preliminary study do not speak in favour of VTs as possible tool in gesture-based AR language rehabilitation. Technology might have in this case detrimental effects on word learning.

**Keywords.** language; gestures; aphasia; rehabilitation, embodiment, virtual training

## 1. Introduction

Language and gesture are two sides of the same coin[3]. When spoken language is impaired, gestures come into play. Patients with anomia (PWA) failing to retrieve single words naming objects or concepts may substitute the words by non-specific words (empty speech), or may provide circumlocutions or gestures[13]. It is conceivable that patients pantomime in order to substitute the words they cannot retrieve. At the same time, patients unconsciously might try to reactivate neural representations linking words and gestures, being both systems processed by a common neural system [15].

---

<sup>1</sup> Manuela Macedonia, Dept. of Information Engineering, Johannes Kepler Universität, Linz, Austria E-mail: [manuela.macedonia@jku.at](mailto:manuela.macedonia@jku.at)

Traditionally, anomia treatment is administered by picture naming through flash cards [5]. In recent years, anomia has also been treated with gestures, thus supporting what patients spontaneously do [2]. This approach finds an early study [4] in which non-fluent aphasics found facilitation in naming objects when performing representative gestures. Despite its potential relevance [11], only a few therapy studies have been conducted on gesture based rehabilitation [12]. In the last few years, digital technologies have paved a new way towards language rehabilitation [6]: with a computer or a tablet, patients can administer themselves as much therapy as they want or need, at any time of the day, ubiquitously [1]. First steps in this direction prove that anomia rehabilitation takes benefit of digital therapy [7]. This option has been tested for naming tasks, with images appearing on the screen. However, digital therapy can be extended to rehabilitation with gestures performed by a virtual trainer in AR.

### *1.2 The present study*

With the present study, we aim to pursue the idea of anomia rehabilitation with gestures by means of a virtual trainer (VT) in augmented reality (AR). We start the project with an experiment on healthy subjects considering that PWA have perception and motion impairments related to their pathology. Here, we hypothesize that imitating the gestures of the avatar is more efficient than hearing and reading the words and watching a VT that performs no gestures.

## **2. Methods**

### *2.1 Participants*

Thirty students German speaking students of the University of Graz (14 F, 16 M; age ranging from 21 to 30;  $M = 26.1$ ,  $SD = 2.88$ ) participated in this study. The study was approved by the local ethics committee.

### *2.2 Stimuli*

The stimuli consisted of 27 items of the artificial corpus “Vimmi” [10]. They were paired with German function words and divided in three counterbalanced learning blocks. For each word, audio-files were recorded. We modelled the AR-Avatar VARA as a woman aged of approximately 40 wearing casual clothes with the editor of game development platform UNITY5 [www.unity.com](http://www.unity.com). The VARA’s skeleton was animated with videos of a human previously recorded by Microsoft Kinect V2, further processed with iPi Studio <http://ipisoft.com/>, Brekel ProBody 2 <https://brekel.com/brekel-pro-body-v2> and Asus Xtion. The avatar performed symbolic gestures that were arbitrarily paired to the words to be memorized. The stimuli consisted therefore of 27 items in Vimmi, 27 audio-files and 27 modelled gestures performed by the avatar. Additionally, a no-gesture sequence, in which the avatar stood still was realized.

### *2.3 Procedure*

In a between-subjects design, participants of Group A, learned 27 artificial words watching the Avatar. The avatar performed a gesture to the word and simultaneously an audio file was played, and the written word appeared on the screen. Thereafter, participants were asked to imitate the avatar’s gesture and to repeat the word aloud (Condition GESTURE / G). In Group B participants performed the same procedure with exception of the gesture: the avatar that staid still performed no gestures and so did

participants while sitting in their chairs (Condition AUDIOVISUAL / AV). Every word was presented 12 times. After each word block there was a 5-minute break. The training lasted three days for approximately one hour daily.

The avatar and the audio files were downloaded into a smartphone (Galaxy S6; Samsung) mounted display on a Google Cardboard (Google) (HMD).

#### 2.4 Tests

After each training participants completed 5 different retention tests on a standard PC by means of Google forms in order to determine their learning progress, i.e. (1) Free recall of German words; (2) Free recall of Vimmi-words; (3) Free recall of German-Vimmi word pairs; (4) Cued recall from German into Vimmi and (5) cued recall from Vimmi into German. Thirty days after the last training, participants were sent a link via email in order to assess their long-term memory performance (Follow-up - FU) with the same tests.

### 3. Results

The correct answers were given a score of 1, and wrong answers were given a score of 0. The score was 0.5 if the answers were not perfect but still recognizable. The scores ranged from 0 to 27 for each test.

The average retrieval performance over all tests over all time points was a mean value of 11.71 (SD=3.78). According to the approach used by Macedonia and colleagues (2011), we further split the Groups in high vs low performers, using the median intra-Group as the reference value (Group A: median=11.82; Group B: median=12). Table 1 reports descriptive data for all the Groups, in all the tests and assessment time-points.

We investigated the influence of gestures on memory performance by running five repeated measures (one for each memory test) ANOVAs with the variable TIME (day 1, day 2, day 3, FU) as within-subject factor, and GROUP (A vs B) as between subject factor. In addition, we considered the two Groups (A and B) separately, and we run five repeated measures (one for each memory test) ANOVAs adding the factor Performance (high vs low) as between subjects variable, in order to understand whether the learning curve differed for high and low performers belonging to the same encoding condition.

In the *first set of analyses*, the factor Time was always significant [Free German:  $F(3,81)= 95.23$ ;  $p<0.001$ ;  $\eta^2=0.78$ ; Free Vimmi:  $F(3,81)= 92.39$ ;  $p<0.001$ ;  $\eta^2=0.77$ ; Paired recall:  $F(3,81)= 85.74$ ;  $p<0.001$ ;  $\eta^2=0.76$ ; German to Vimmi:  $F(3,81)= 87.49$ ;  $p<0.001$ ;  $\eta^2=0.76$ ; Vimmi to German:  $F(3,36)= 68.43$ ;  $p<0.001$ ;  $\eta^2=0.71$ ]; repeated contrasts indicated differences from T1 and T2, from T2 and T3, and from T3 and T4 in all the tests (with Bonferroni correction for multiple comparisons all  $p(s) <0.05$ ). The factor Group was significant only in the German to Vimmi Test ( $F(1,27)= 4.59$ ;  $p=0.04$ ;  $\eta^2=0.14$ ], indicating that participants in Group B learned more than those in Group A (mean A= 8.97; mean B= 12). None of the interactions Time X Group was significant. Therefore, we conclude that the gestures did not affect the learning curve differently from the still condition.

In the *second set of analyses* we considered Group A and B separately. For Group A, the factor Time resulted significant in all the memory tests [Free German:  $F(3,39)= 47.87$ ;  $p<0.001$ ;  $\eta^2=0.79$ ; Free Vimmi:  $F(3,39)= 46.64$ ;  $p<0.001$ ;  $\eta^2=0.78$ ; Paired recall:  $F(3,39)= 47.75$ ;  $p<0.001$ ;  $\eta^2=0.79$ ; German to Vimmi:  $F(3,39)= 46.38$ ;  $p<0.001$ ;  $\eta^2=0.78$ ; Vimmi to German:  $F(3,39)= 37.78$ ;  $p<0.001$ ;  $\eta^2=0.74$ ]; repeated contrasts

indicated differences from T1 and T2, from T2 and T3, and from T3 and T4 in all the tests (with Bonferroni correction for multiple comparisons all  $p(s) < 0.05$ ).

Not surprisingly, high performers in general learned more than low performers [Free German:  $F(1,13) = 24.26$ ;  $p < 0.001$ ;  $\eta^2 = 0.65$ ; Free Vimmi:  $F(1,13) = 17.01$ ;  $p < 0.001$ ;  $\eta^2 = 0.57$ ; Paired recall:  $F(1,13) = 29.78$ ;  $p < 0.001$ ;  $\eta^2 = 0.7$ ; German to Vimmi:  $F(1,13) = 24.22$ ;  $p < 0.001$ ;  $\eta^2 = 0.65$ ; Vimmi to German:  $F(1,13) = 14.97$ ;  $p = 0.02$ ;  $\eta^2 = 0.53$ ]. More interestingly, the interaction Time X Performance was significant only in the Paired recall [ $F(3,39) = 5.55$ ;  $p = 0.03$ ;  $\eta^2 = 0.3$ ] and in the German to Vimmi test [ $F(3,39) = 3.4$ ;  $p = 0.03$ ;  $\eta^2 = 0.21$ ]. A closer look at the differences among the levels of the interaction evidenced that in the Paired recall the high performers learned more than the low performers in T2 compared to T1 [ $F(1,13) = 8.63$ ;  $p < 0.05$ ;  $\eta^2 = 0.4$ ] but also their performance decreased more than that of the low performers from T3 to T4 [ $F(1,13) = 12.94$ ;  $p < 0.05$ ;  $\eta^2 = 0.5$ ]. In the German to Vimmi test high performers lost more of the acquired words than low performers from T3 to T4 [ $F(1,13) = 7.9$ ;  $p < 0.05$ ;  $\eta^2 = 0.38$ ] (all the comparisons were corrected with Bonferroni correction for multiple comparisons). Figure 2 illustrates these interaction effects. For the Group B, both main effects of Time [Free German:  $F(3,36) = 44.12$ ;  $p < 0.001$ ;  $\eta^2 = 0.79$ ; Free Vimmi:  $F(3,36) = 50.67$ ;  $p < 0.001$ ;  $\eta^2 = 0.81$ ; Paired recall:  $F(3,36) = 52.77$ ;  $p < 0.001$ ;  $\eta^2 = 0.82$ ; German to Vimmi:  $F(3,36) = 51.84$ ;  $p < 0.001$ ;  $\eta^2 = 0.81$ ; Vimmi to German:  $F(3,36) = 31.6$ ;  $p < 0.001$ ;  $\eta^2 = 0.73$ ] and Performance [Free German:  $F(1,12) = 7.82$ ;  $p < 0.001$ ;  $\eta^2 = 0.97$ ; Free Vimmi:  $F(1,12) = 7.14$ ;  $p = 0.02$ ;  $\eta^2 = 0.37$ ; Paired recall:  $F(1,12) = 1664$ ;  $p = 0.02$ ;  $\eta^2 = 0.58$ ; German to Vimmi:  $F(1,12) = 14.28$ ;  $p = 0.03$ ;  $\eta^2 = 0.54$ ; Vimmi to German:  $F(1,12) = 14.97$ ;  $p = 0.02$ ;  $\eta^2 = 0.55$ ] were found. Repeated contrasts on the different levels of Time underlined differences statistically significant ( $p < 0.05$ ) from T1 to T2, from T2 to T3, and from T3 to T4 (with Bonferroni correction). However, none of the interactions Time X Performance reached significance, indicating that the learning curve did not differ in high and low performers who learned words in still condition.

#### 4. Discussion

In this preliminary study, we tested a rehabilitation scenario with healthy subjects in order to assess for the feasibility of embodied learning by means of a VT in AR. Briefly, gestural training compared to audio-visual training showed no memory enhancement for words. When splitting the two training groups in high and low performers, in Group A, high performers benefitted more of gestures than low performers in the recall test from German into Vimmi and in the cued paired recall.

The overall results do not match the hypothesis, i.e., gesture training is more effective than audio-visual training. The reasons leading to this poor result can be multiple. First, a between subject design is a limitation: subjects might display different cognitive capacities in both groups. Second, the duration of the training may not have been sufficient. Third, these results may be attributed to the use of technology.

Considering the interactions between level of performance and training, the present data do not match the results of another study conducted with the same vocabulary items and similar gestures [9]. The present study provides rather evidence for the Theory of Cognitive Load (TCL). It describes the limits of our cognitive processing capacities with focus on our working memory. Thereafter multimodal input would enhance mental workload and thus penalize low performers in memory tasks [14]. We conclude that by the moment that PWA should not take the burden of AR training conducted with a VT

in AR.

## References

- [1] M. Agostini, M. Garzon, S. Benavides-Varela, S. De Pellegrin, G. Bencini, G. Rossi, S. Rosadoni, M. Mancuso, A. Turolla, F. Meneghello, and P. Tonin, Telerehabilitation in Poststroke Anomia, *Biomed Res. Int.* **2014** (2014) 1–6. doi:10.1155/2014/706909.
- [2] L. Dipper, M. Pritchard, G. Morgan, and N. Cocks, The language–gesture connection: Evidence from aphasia, *Clin. Linguist. Phon.* **29** (2015) 748–763. doi:10.3109/02699206.2015.1036462.
- [3] S. Goldin-Meadow, and D. Brentari, Gesture and language: Distinct subsystem of an integrated whole, *Behav. Brain Sci.* **40** (2017) e74. doi:10.1017/S0140525X1600039X.
- [4] R.E. Hanlon, J.W. Brown, and L.J. Gerstman, Enhancement of naming in nonfluent aphasia through gesture., *Brain Lang.* **38** (1990) 298–314. <http://www.ncbi.nlm.nih.gov/pubmed/2322814> (accessed July 3, 2019).
- [5] S.E. Kohn, and H. Goodglass, Picture-naming in aphasia., *Brain Lang.* **24** (1985) 266–83. <http://www.ncbi.nlm.nih.gov/pubmed/3978406> (accessed July 3, 2019).
- [6] M. Lavoie, J. Macoir, and N. Bier, Effectiveness of technologies in the treatment of post-stroke anomia: A systematic review, *J. Commun. Disord.* **65** (2017) 43–53. doi:10.1016/j.jcomdis.2017.01.001.
- [7] M. Lavoie, S. Routhier, A. Légaré, and J. Macoir, Treatment of verb anomia in aphasia: efficacy of self-administered therapy using a smart tablet, *Neurocase.* **22** (2016) 109–118. doi:10.1080/13554794.2015.1051055.
- [8] M. Macedonia, and K. Mueller, Exploring the neural representation of novel words learned through enactment in a word recognition task, *Front. Psychol.* **7** (2016). doi:10.3389/fpsyg.2016.00953.
- [9] M. Macedonia, K. Müller, and A.D. Friederici, Neural correlates of high performance in foreign language vocabulary learning, *Mind, Brain, Educ.* **4** (2010). doi:10.1111/j.1751-228X.2010.01091.x.
- [10] M. Macedonia, K. Müller, and A.D. Friederici, The impact of iconic gestures on foreign language word learning and its neural substrate, *Hum. Brain Mapp.* **32** (2011) 982–998. doi:10.1002/hbm.21084.
- [11] J. Marshall, The roles of gesture in aphasia therapy, *Adv. Speech Lang. Pathol.* **8** (2006) 110–114. doi:10.1080/14417040600672384.
- [12] J. Marshall, A. Roper, J. Galliers, S. Wilson, N. Cocks, S. Muscroft, and T. Pring, Computer delivery of gesture therapy for people with severe aphasia, *Aphasiology.* **27** (2013) 1128–1146. doi:10.1080/02687038.2013.786803.
- [13] M.L. Rose, Z. Mok, and K. Sekine, Communicative effectiveness of pantomime gesture in people with aphasia, *Int. J. Lang. Commun. Disord.* **52** (2017) 227–237. doi:10.1111/1460-6984.12268.
- [14] J. Sweller, Cognitive Load During Problem Solving: Effects on Learning, *Cogn. Sci.* **12** (1988) 257–285. doi:10.1207/s15516709cog1202\_4.
- [15] J. Xu, P.J. Gannon, K. Emmorey, J.F. Smith, and A.R. Braun, Symbolic gestures and spoken language are processed by a common neural system, *Proc. Natl. Acad. Sci.* **106** (2009) 20664–20669. doi:10.1073/pnas.0909197106.