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

Manuela MACEDONIA[a,1], Joachim GREINER [c], Claudia REPETTO [d],
Selina C. WRIESSNEGGER [c]
[a] *Department of Information Engineering, Johannes Kepler University Linz, Austria*
[b] *Max-Planck Institute for Human Cognitive and Brain Sciences Leipzig, Germany*
[c] *Institute of Neural Engineering, Technical University Graz, Austria*
[d] *Department of Psychology, Catholic University of Sacred Heart, 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, or 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 on 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 word 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 them. Participants were trained singularly for 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 from gesture-based training than low performers. The data in this preliminary study do not speak in favour of VTs as possible tools 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 [1]. When spoken language is impaired, gestures come into play. Patients with anomia (PWA) failing to retrieve single words when naming objects or concepts may substitute the words with non-specific words (empty speech), or may provide circumlocutions or gestures [2]. It is conceivable that patients pantomime in order to substitute the words they cannot retrieve. At the same time, patients might unconsciously try to reactivate neural representations linking words and gestures, being as both systems are processed by a common neural system [3]. Traditionally, anomia treatment is administered by picture naming through flash cards [4]. In recent years, anomia has also been treated with gestures, thus supporting what patients spontaneously do [5]. This approach finds an early study [6] in which non-fluent aphasics found facilitation in naming objects when performing representative gestures. Despite its potential relevance [7], only a few therapy studies have been conducted on gesture-based rehabilitation [8]. In the last few years, digital technologies have paved a new way towards language rehabilitation [9]: 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 [10].

[1] Corresponding Author: manuela@macedonia.at.

First steps in this direction prove that anomia rehabilitation takes benefit of digital therapy [11]. 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 will 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 German-speaking students from 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” [12]. 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). 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 for the word. 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 the exception of the gesture. The avatar that remained 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: (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

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 [13], 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 the within-subject factor, and GROUP (A vs B) as the between-subject factor. In addition, we considered the two Groups (A and B) separately, and we ran five repeated measures (one for each memory test) ANOVAs adding the factor Performance (high vs low) as the 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; η^2=0.78; Free Vimmi: F(3,81)= 92.39; p<0.001; η^2=0.77; Paired recall: F(3,81)= 85.74; p<0.001; η^2=0.76; German to Vimmi: F(3,81)= 87.49; p<0.001; η^2=0.76; Vimmi to German: F(3,36)= 68.43; p<0.001; η^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; η^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 of Time X Group were 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 was significant in all the memory tests [Free German: F(3,39)= 47.87; p<0.001; η^2=0.79; Free Vimmi: F(3,39)= 46.64; p<0.001; η^2=0.78; Paired recall: F(3,39)= 47.75; p<0.001; η^2=0.79; German to Vimmi: F(3,39)= 46.38; p<0.001; η^2=0.78; Vimmi to German: F(3,39)= 37.78; p<0.001; η^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; η^2=0.65; Free Vimmi: F(1.13)= 17.01; p<0.001; η^2=0.57; Paired recall: F(1,13)= 29.78; p<0.001; η^2=0.7; German to Vimmi: F(1,13)= 24.22; p<0.001; η^2=0.65; Vimmi to German: F(1,13)= 14.97; p=0.02; η^2=0.53]. More interestingly, the interaction of Time X Performance was significant only in the Paired recall [F(3,39)= 5.55; p=0.03; η^2=0.3] and in the German to Vimmi test [F(3,39)= 3.4; p=0.03; η^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; η^2=0.4] but their performance also decreased more than that of the low performers from T3 to T4 [F(1,13)= 12.94; p<0.05; η^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; η^2=0.38] (all the comparisons were corrected with Bonferroni correction for multiple comparisons). Figure 2 illustrates these interaction effects. For Group B, both main effects of Time [Free German: F(3,36)= 44.12; p<0.001; η^2=0.79; Free Vimmi: F(3,36)=50.67; p<0.001; η^2=0.81; Paired recall: F(3,36)= 52.77; p<0.001; η^2=0.82; German to Vimmi: F(3,36)= 51.84; p<0.001; η^2=0.81; Vimmi to German: F(3,36)= 31.6; p<0.001; η^2=0.73] and Performance [Free German: F(1,12)= 7.82; p<0.001; η^2=0.97; Free Vimmi: F(1,12)= 7.14; p=0.02; η^2=0.37; Paired recall: F(1,12)= 1664; p=0.02; η^2=0.58; German to Vimmi: F(1,12)= 14.28; p=0.03; η^2=0.54; Vimmi to German: F(1,12)= 14.97; p=0.02; η^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 of Time X Performance reached significance, indicating that the learning curve did not differ between high and low performers who learned words in the 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 from 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 [14]. The present study rather provides 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 [15]. We conclude that, for the moment, PWA should not take on the burden of training conducted with a VT in AR.

References

[1] Goldin-Meadow S, Brentari D. Gesture and language: Distinct subsystem of an integrated whole. Behav Brain Sci. 2017 Apr 26;40:e74.

[2] Rose ML, Mok Z, Sekine K. Communicative effectiveness of pantomime gesture in people with aphasia. Int J Lang Commun Disord. 2017 Mar;52(2):227–37.

[3] Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR. Symbolic gestures and spoken language are processed by a common neural system. Proc Natl Acad Sci. 2009 Dec 8;106(49):20664–9.

[4] Kohn SE, Goodglass H. Picture-naming in aphasia. Brain Lang. 1985 Mar;24(2):266–83.

[5] Dipper L, Pritchard M, Morgan G, Cocks N. The language–gesture connection: Evidence fromaphasia. Clin Linguist Phon. 2015 Oct 3;29(8–10):748–63.

[6] Hanlon RE, Brown JW, Gerstman LJ. Enhancement of naming in nonfluent aphasia through gesture. Brain Lang. 1990 Feb;38(2):298–314.

[7] Marshall J. The roles of gesture in aphasia therapy. Adv Speech Lang Pathol. 2006 Jan 3;8(2):110–4.

[8] Marshall J, Roper A, Galliers J, Wilson S, Cocks N, Muscroft S, et al. Computer delivery of gesture therapy for people with severe aphasia. Aphasiology. 2013 Sep;27(9):1128–46.

[9] Lavoie M, Macoir J, Bier N. Effectiveness of technologies in the treatment of post-stroke anomia: A systematic review. J Commun Disord. 2017 Jan;65:43–53.

[10] Agostini M, Garzon M, Benavides-Varela S, De Pellegrin S, Bencini G, Rossi G, et al.Telerehabilitation in Poststroke Anomia. Biomed Res Int. 2014;2014:1–6.

[11] Lavoie M, Routhier S, Légaré A, Macoir J. Treatment of verb anomia in aphasia: efficacy of self-administered therapy using a smart tablet. Neurocase. 2016 Jan 2;22(1):109–18.

[12] Macedonia M, Müller K, Friederici AD. The impact of iconic gestures on foreign language word learning and its neural substrate. Hum Brain Mapp. 2011 Jun 1;32(6):982–98.

[13] Macedonia M, Mueller K. Exploring the neural representation of novel words learned through enactment in a word recognition task. Front Psychol. 2016;7s.

[14] Macedonia M, Müller K, Friederici AD. Neural correlates of high performance in foreign languagevocabulary learning. Mind, Brain, Educ. 2010;4(3).

[15] Sweller J. Cognitive Load During Problem Solving: Effects on Learning. Cogn Sci. 1988 Apr 1;12(2):257–85.