Joachim Greiner, BSc.

Foreign language acquisition with gestures in an augmented reality environment

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Biomedical Engineering

submitted to

Graz University of Technology

Supervisor

Mag. Dr. Selina C. Wriessnegger,

Institute of Neural Engineering

Univ.-Prof. Dipl.-Ing. Dr.techn. Müller-Putz
Head of Institute

Graz, September 2018

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

Date

Signature

Acknowledgements

First, I would like to thank Mag. Dr. Selina Christin Wriessnegger for supervising this thesis and for her continuous expert advice, encouragement and patience during its progress, as well as Assoc. Prof. Dipl. Ing. Dr. techn. Gernot Müller-Putz for enabling this thesis.

My gratitude is also owed to the participants in my study, who dutifully fulfilled their part in their spare time.

Nobody has been more important to me in my life than the members of my family. I would like to thank my parents, who supported me with their generosity, guidance and love. Most importantly, I wish to thank my loving and supportive partner, Heike Brunner who provided unending inspiration.

Zusammenfassung

Das Erlernen von Fremdsprachen ist ein elementarer Bestandteil einer jeden fundierten Bildung und liefert unzählige Vorteile im Leben. Eine wichtige Komponente dabei ist das Einprägen einzelner Wörter und deren Übersetzung. Der positive Einfluss von Gestik auf das Erlernen von konkreten Wörtern wurde bereits mehrfach bestätigt. Ziel dieser Arbeit ist es zu zeigen, dass dies auch für abstrakte Wörter gilt.

In dieser Studie wurde der Einfluss von Gestik während des Erlernens von abstrakten Fremdwörtern durch Unterstützung von augmented reality (AR) und eines Virtual Agents (VARA) untersucht. Zusätzlich wurde das Elektroenzephalogramm (EEG) während des Übersetzens aufgezeichnet. In drei Lernphasen wurden 27 Kunstwörter (Vimmi) zusammen mit deutschen abstrakten Wörtern als Übersetzung erlernt. Die Studie wurde in einem „between-subject“ Design konzipiert, bei der 15 ProbandInnen mit und 15 ProbandInnen ohne Gestik lernten. Recall Tests wurden durchgeführt, um den Lernfortschritt zu überprüfen. Die EEG Messung wurde nach den drei Lerneinheiten durchgeführt, um Unterschiede zwischen der Gestik und der Nicht-Gestik Gruppe in Ereigniskorrelierten Potentialen (ERPs) sichtbar zu machen.

Die Lerngruppen wurden mittels Mediansplit anhand ihres Lernerfolges in „High-“ und „Low-Performers“ unterteilt. Es konnte kein signifikanter Unterschied zwischen den High-Performern, sowie zwischen den Low-Performern beider Gruppen festgestellt werden. Obwohl die Gestik-Gruppe mit einem Nachteil startete, konnte sie zu der Nicht-Gestik-Gruppe aufschließen und hätte diese womöglich übertreffen, wenn es weitere Lerneinheiten gegeben hätte. Ähnlich verhielt es sich bei den ERP Ergebnissen. Amplituden der N200 sowie N400 Komponenten waren niedriger für die Nicht-Gestik-Gruppe, aber dieser Unterschied war nicht signifikant.

Die Resultate sind ein weiterer Hinweis dafür, dass Gestik eine geeignete Lernunterstützung für Fremdwörter darstellt. Es gibt keinen negativen Einfluss bei Anwendung von Gesten beim Erlernen von abstrakten Wörtern in wenigen Einheiten und einen steigenden Vorteil je öfter gelernt wird.

Stichwörter: Spracherwerb, Enactment, EEG, ERP, N200, N400, Augmented Reality, Virtual Agent

Abstract

The acquisition of a foreign language is an elementary building block of every fundamental education and provides innumerable advantages in life. An important part in this process is memorization of single words and their translation. The positive influence of gestures during learning of concrete words was already demonstrated multiple times. The goal of this thesis is to show that this is also true for abstract words.

In this study the influence of gestures during foreign word acquisition by means of augmented reality and a virtual agent (VARA) is examined. Additionally the electroencephalogram (EEG) was recorded to investigate brain activity during translation. In three learning phases 27 artificial words (Vimmi), together with abstract German words as their translation, had to be learned. The study was drafted with a between-subject design, with 15 subjects learning with and 15 subjects without the aid of gestures. Recall tests were conducted to review the memory performance of each subject. The EEG measurement was carried out after the three learning units, to determine differences in ERP components between gesture and no gesture group.

Based on their memory performance the groups were further divided into high and low performers by a median split. No significant difference between the high and the low performers of both groups could be found. But the gesture groups, although being at a disadvantage in the beginning, were able to catch up to the no gesture groups on day 3, indicating, that they would surpass them with additional training units. Similar outcomes were obtained for ERPs. N200 and N400 ERP component amplitudes were lower for the no gesture group, but it was not a significant difference.

These results indicate that gestures are a useful learning tool for foreign language acquisition. There is no negative influence of using gestures, even when learning abstract words over a short period of time and increased benefits for additional learning units.

Key words: Foreign language learning, Enactment, EEG, ERP, N200, N400, Augmented Reality, Virtual Agent

Content

List of Tables

List of Figures

1. Introduction

The following chapter highlights principles that are the basis of this work. In section 1.1 I will introduce language learning with gestures, followed by an overview of the support augmented reality and a virtual agent can offer in section 1.2. Section 1.3 will offer an insight into Electroencephalography (EEG) and Event Related Potentials (ERP). Following these descriptions the goals of this thesis are explained.

1.1. Language learning with gestures

Learning foreign languages is a fundamental building block of every modern education and is therefore implemented as early as primary school, or even earlier if parents raise their children bilingual. The benefits of knowing one or more foreign languages are self-evident, and stretch from advantages in school, social life or professional field, to simply being able to enjoy foreign multimedia content.

Language evolved from signs combined with vocalizations, so gestures predate words in human communication. When learning a novel word through a combination of gesture and language, two parts of an ancient communicative system are activated [6].

Foreign language learning in the 21st century mainly occurs via reading and listening, for example through texts with gaps, or by repeating bilingual lists [2]. From as early as 1976, when Piaget described the native language acquisition as a sensorimotor process [3], to today, when motor acts and multisensory perception is demonstrated to be involved in language learning, by cognitive sciences [4,5], we see that there is great potential in foreign language learning with the use of gestures, which currently is not fully utilized. Gestures performed during learning of vocabulary words and phrases can enhance memory [7], improve foreign language learning [8, 9, and 10] and delay their unlearning.

The earliest report of a positive effect of gestures dates back to 1768. Radonvilliers noted that native language (L1) is learned with pictures and gestures, but in foreign language learning (L2) this is not the case anymore [11]. In 1969 the Total Physical

1. Introduction

Response approach was observed: Performing actions from L2 commands led to better memory of command phrases [12]. Action words (for example “cut”) in L1 are memorized better when subjects performed gestures. The term “enactment effect” for this phenomenon was coined by Engelkamp in 1980 [13]. Similar studies with L1 around the same time (1981) named it the “subject performed task effect” [14].

The first empirical study of the influence of gestures on memory for L2 was conducted in 1995. English natives learned French sentences with and without gestures. Enacted phrases elicited better short- and long-term memory results [15]. Subjects were able to understand more detailed information when speakers produced iconic gestures with their speech (typing gesture when describing writing) [16]. In a study by Macedonia in 2003, German speaking university students learned words from an artificial corpus audio-visual and with the aid of gestures. Words were memorized significantly better when gestures were incorporated in the learning process [9].

With the help of fMRI, it was demonstrated that gestural and spoken information are processed in a comparable fashion during sentence comprehension in the Borca’s area, as described by Willems (2007) [17, 18].

Tellier (2008) taught French preschoolers English words, either with pictures or self-performed iconic gestures. Learning with gestures yielded higher memorization results [19]. A year later English natives learned Japanese verbs audio-visually and by performing an iconic gesture that was either incongruent or congruent. The latter led to better results, while the group with incongruent gestures scored lowest [20]. Macedonia (2011) demonstrated that iconic gestures yielded better short- and long-term results when comparing illustrative and meaningless gestures while learning words [21].

Furthermore the enactment effect was tested with abstract words learned, embedded in sentences, with gestures that showed enhancement in memory. For abstract words, gestures create an arbitrary new motor image that grounds abstract meaning in the learner’s body [22].

Bergmann and Macedonia (2013) compared an anthropomorphic virtual agent and a human trainer. Subjects learned L2 words by reading and hearing, or by performing gestures as well. With both trainers the results were better when learning with gestures [23]. Mayer (2014) taught novel words of an artificial corpus either with a picture or a gesture (iconic gesture, or drawing of the outline of the concept). Words learned with iconic gestures scored better [24]. A 2014 study to test enactment demonstrated that self-performance of gestures enhances learning. School children and young adults learned L2 words with the aid of a virtual agent that demonstrated the gestures [25, 26].

These findings show that by performing an iconic gesture during learning of novel words of a foreign language, the ability to memorize and retain that information over longer periods of time is greatly increased. Furthermore this effect is more substantial with gestures that represent the word's semantics or some features of it, than with meaningless gestures.

1.1.1. Learning theories

There are multiple factors that explain the enactment effect: Creation of memory representation, interconnectedness of language and gestures and the relationships between them [1]. This increases the words representation in the mind. If one part of the representation decays, another can make the word accessible again.

A) Memory Trace:

The increased memorization by execution of gestures is due to the creation of a motor trace [27, 28]. A neuroscientific study documented, that words learned with the help of gestures evoke activity in the motor cortex [29, 21]. It was also attributed to the complexity of the memory trace [19, 30].

B) Depth of encoding:

This concept proposed different levels of information processing [31]. By adding a gesture and with it sensorimotor information, the code for the word grows more complex and therefore improves retrieval in memory [34, 35]. Only hearing a word is encoded shallowly and adding a picture or gesture deepens the encoding. Findings of multiple studies were explained by depth of encoding [15, 19, 20, 21, 32, 33].

C) Mental Imagery:

Subjects that act out gestures during learning activate an internal kinetic image of the word’s semantic [20, 21, 22]. Gestures emerge from an underlying mental image of concepts and therefore are tightly connected to them [36]. Therefore a person can access a word in L2 not only from the word in L1, but also through this mental image.

D) Supra-modal semantic network:

In the brain the supra-modal semantic network serves both speech and gesture semantics [37]. An overlapping network of brain regions is activated, when information from speech and iconic gestures are processed

E) Enhanced Attention

Gestures enhance the attention of subjects when compared to groups that use simpler techniques, such as bilingual lists [6, 98, 99, 58, 100]. This induces representational stability in the hippocampus [58, 59].

Other underlying effects and theories include that words are experience dependent functional networks. For example reading the word cinnamon activates the network with sensory information for smell, taste, touch, look and motor programs for interaction. Even if no cinnamon is present, the brain’s regions processing odor and taste are activated as well [38, 39, 40].

If subjects enact a gesture during learning of a novel foreign word, they have a multimodal sensorimotor experience, during which sensorimotor patterns, previously constructed in native language learning, are reproduced and reinforced. This effect is more prominent in action words or phrases, such as to go or to give [6, 41].

The concreteness effect occurs when concrete words in combination with iconic gestures are used, due to easier processability [42] and better memorability [43, 44, 45].

In the case of abstract words, arbitrary gestures have to be used. Gestures can make the concept of the words more concrete (concreteness effect [22].). Adverbs have low representation as images and do not possess strong sensorimotor representations [60, 61, 62]. But they are linked to socio-linguistic information (Words as Social Tools Theory [63, 64]) and they possess an emotional valence [65, 66]. Arbitrary gestures can create a new motor image, which extends the original representation of the adverb with a sensorimotor component. This in turn leads to increased retention over adverbs learned without any gestures [46]. In a 2017 study the effect of enrichment training for abstract words was examined. Words learned with the aid of gestures were better remembered in free L1 recall tests and were also better recognized [95].

Figure 1 shows the word network for cinnamon in the brain. Besides canonical language areas, the network contains areas processing and storing sensorial information that was experienced in combination with cinnamon (visual, odor, taste, color, texture and so on). It is also connected to motor areas used for actions with cinnamon.

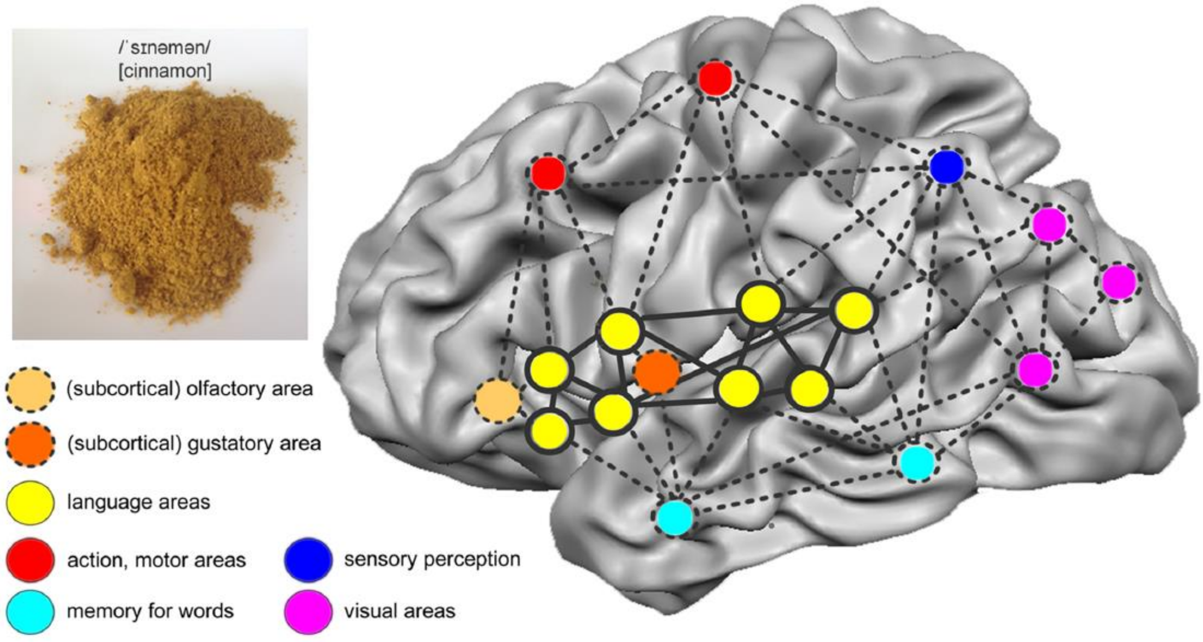


Figure 1: Word network for cinnamon [1]

1.2. Augmented Reality

One of the most prominent uses of augmented reality (AR) in recent years was Pokémon GO. It allows players to catch Pokémon in AR that can be found in different locations in the real world. It was linked to language learning through the aspect of creating a personal avatar, with vocabulary describing sex, dress, eye and hair color [103].

The current generation is often described as digital native and utilizing different types of technology can make language learning more inspiring, motivating and more meaningful [104]. Incorporating AR technology makes learning more effective and the results are longer lasting. Standard 2D practices are improved by enhanced 3D content [105] and by doing so it bridges the gap between virtual and real world [106]. One of the most efficient methods to improve foreign language skills is visiting communities, where the desired language is spoken [107.] AR can help to simulate these experiences in a classroom environment.

Many AR applications with a focus on language learning rely on optical sensors (digital cameras for computers or handheld devices) in combination with AR markers (quick response (QR) codes, or black and white square printed objects). They are used to trigger actions in the application and present different types of content. Markers are easy to process, which makes AR available for a wide variety of devices [108].

Advantages of AR learning tools include increased motivation amongst students, emotional engagement and the possibility of direct feedback to learners. Further involvement of students can be achieved by student-generated AR. Pupils are part of content creation, which is linked to improved learning outcomes [109, 110].

Further positive influences are that abstract concepts can be concretized [111] and therefore better understood [112], improved critical thinking and problem solving [113], improved comprehension [114] as well as the possibility to accommodate different learning styles [115].

Projects in which marker-based AR was used by language educators included English to Tamil translation [116], learning Filipino and German vocabulary [117] or pronunciation of English words [118]. A location based AR app is used as a campus tour that additionally teaches English [119]. The app Mentira teaches Spanish pragmatics inside and outside of the classroom as well as in independent gameplay by letting pupils solve a prohibition-era murder [120].

Audio-visual content increase motivation by providing better learning environment for primary school pupils which improve learning [121], this is also shown by a study using an AR popup book [122]. Participants in a study that implemented augmented reality in language learning showed improved motivational levels towards vocabulary learning [118].

1.2.1. Virtual Agents

A way to incorporate gestures into novel word or language learning is with the help of a virtual agent in an augmented reality environment. They are essentially interfaces of learning programs that have an anthropomorphic appearance. They can articulate the words that have to be learned and perform accompanying gestures like a human being.

The agent would allow students to work on their language skills self-reliant without the need of an instructor or teacher. This allows them to freely choose a time and a place of learning. Such an agent is also required, since the transfer of information is multimodal and includes the written word, audio reproduction and the visualization of gestures. They are also easily accessible since learning apps with a virtual agent can be installed on almost any smartphone and they can incorporate virtual or augmented reality as well.

Other advantages of the presence of virtual agents can be increased motivation [47, 48], social interactions (nodding or facial expressions) by the agent [49], they can direct learners’ attention [50] and offer support to learners [51]. The goal is to make them more lifelike [52] for a more positive effect on learners through the personal effect [53].

A number of studies have been conducted, to see whether a virtual agent has a positive influence on subjects compared to human trainers or no trainer at all:

For example, Billie (see Figure 2) successfully trained humans on vocabulary learning by means of enactment in 2011 [54]. He was also utilized and compared to a human trainer with adult subjects. Both groups trained equally well, but high performers reached better results with the help of Billie [23]. Adults rated the agent's (Billie's) gesture quality and personality compared to a human trainer (adult female) as less natural, but there was no significant difference in their personality score [55]. Children were asked to rate a human (12 year old boy) and virtual trainer (Billie) on gestures and personalities. The human trainer garnered more sympathy and naturalness, but the difference was not significant [56]. In another study Billie was used to test enactment with school children in a classroom environment. This proved to be successful and repeating gestures yielded a better learning payoff [57].

Figure 2 Billie, a virtual agent representing a 11 or 12 year old boy, performing the gesture for "mug" (screenshots from the video) [57].

1.3. EEG

Hans Berger was the first person to measure human EEG in experiments that demonstrated the possibility to measure electrical activity of the brain. He placed clay electrodes onto the scalp and was able to plot continuous oscillations after large magnification [67]. Over time a number of similar ongoing EEG brain activities were found, which are not driven by discrete events. They are feedback loops in the brain that are oscillatory and are classified by their frequency bands [96]:

- Delta band < 4 Hz
- Theta band 4 – 8 Hz
- Alpha band 8 – 13 Hz
- Beta band 13 – 30 Hz
- Gamma band > 30 Hz

For example the alpha wave oscillation with approximately 10 Hz has the strongest amplitude over the back of the head when the subject's eyes are closed. When working with a larger number of trials, the alpha oscillation will average close to zero.

The number of electrodes depends on the experiment and can span from only a few up to 256 across the head. The most common system for electrode positioning is the International 10 – 20 System [68, 69]. The lines between Nasion Nz and Inion Iz (distinct locations on the front and back of the skull) and the left and right pre-auricular points (depression behind the ears) are split at 10% or 20% points. Electrodes are then equally distributed along the arcs between them. The names for the positions correspond with their location (Fp = frontal pole; F = frontal; C = central; P = parietal; O = occipital; and T = temporal) and end with odd or even numbers for the left or right hemisphere, starting from the middle line with z for zero (Figure 3).

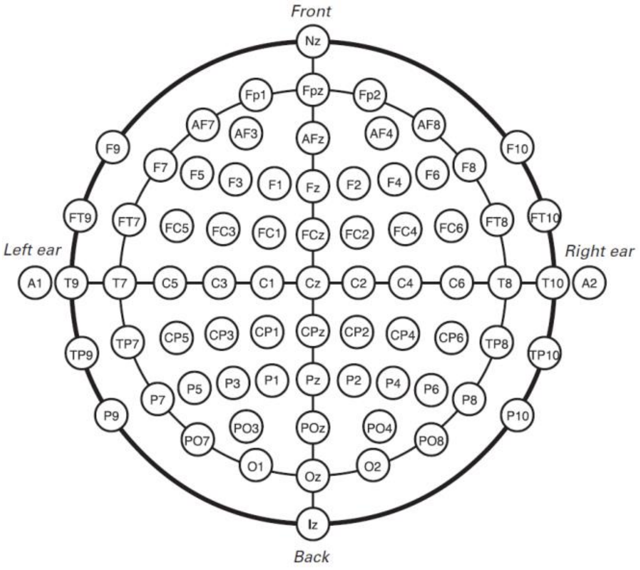


Figure 3: International 10/20 System for electrode placement [68, 69]

Since signals from the electrodes are in the range of microvolts, they have to be amplified by a factor of 1000 – 100.000. A good electrical connection between electrodes and scalp has to be ensured by measuring the electrode impedance and reducing it by adding electrode gel. Afterwards the continuous signal of each channel is sampled at a rate of 200 to 1000 to create discrete digital values [96].

1.3.1. Preprocessing

Another part of ERP analysis is baseline correction. It works on the assumption that the voltage during the baseline period is a good estimate of the voltage offset for the rest of the trial. The average voltage of the baseline is then removed from the epoch, eliminating this offset. [96]

There are a number of different artifacts that superimpose the signals we want to analyze. Most common are eye movement, or blinking, muscle artifacts and potentials from outside sources [96]. Trials that are corrupted by these are usually rejected or filters are used to suppress noise. For example, to reduce electrode drift (very slow voltage change) a high pass with a low cut off frequency (<0.1 Hz) can be applied. High frequency noise can be removed by using a low pass with a cut off frequency above the desired signals frequency (> 15-100 Hz). To prevent distortion of the ERPs high pass filters higher than 0.5 Hz and low pass filters smaller than 10 Hz should not be implemented [96].

1.3.2. ERPs

EEG is a rough measurement since it represents the summation of many different neural sources of activity within the brain. To extract feedback of the brain to certain events, the EEG signal has to be subjected to an averaging technique, to create event related potentials (ERPs). They mostly originate as postsynaptic potentials (PSPs), which occur when the flow of ions across cell membranes changes due to neurotransmitters. If PSPs from thousands of neurons, that are aligned similarly (mostly pyramidal cells of the cerebral cortex), arise at the same time, they summate and conduct through the brain to the scalp, where they can be measured. When PSPs originate at the end of cortical pyramidal neurons, they form dipoles and depending on the receiving electrode's location, the resulting ERP signal can be positive or negative. This can also be seen in the name of the ERP components (positive: P300, P600, or negative: N400). [96]

Since ERPs are very small compared to the rest of the EEG activity of the brain, they are averaged over multiple trials. To do so, time-locking points are needed that are stored as event codes in addition to the EEG. This is the case e.g. at every onset of each stimulus or at the press of a button. The time before the stimulus is called pre-stimulus baseline period and the time after it shows the ERPs. Dividing the EEG into these fixed length trial segments is called EPOCHING. Each of these epochs look different from one another. This is due to the fact, that EEG is a sum of many different electrical sources within the brain. The EEG segments corresponding to one condition of an experiment are averaged together into one waveform. If enough trials are used, activities that did not originate because of the stimulus will cancel each other out, due to their random nature [96].

Contingent negative variation (CNV) was the first ERP to be discovered [70]. Subjects were told to press a button when a target flash, which was preceded by a warning signal, appeared. The CNV was found between warning and target, indicating the preparation for the task.

In 1965 the P3 component was uncovered [71]. Subjects were presented with auditory or visual stimuli that lead to a positive peak at about 300 ms post stimulus. The P3 was larger, when subjects did not know which stimuli they would be presented with.

After these initial findings researches focused on identifying various cognitive ERP components.

When it comes to language, ERP components related with basic operations (phoneme discrimination; word segmentation) are early (100 – 200 ms), fast and automatic, while other components that describe integration or revision processes tend to have latencies of up to 1000 ms [97].

N100 word segment processing: Word onset syllables cause larger anterior N100 responses than word medial syllables. Especially stressed syllables caused larger N100 components than unstressed syllables closer to the midline [72].

Another ERP component is the N170, which shows that the human brain is capable of differentiating between faces and other objects as soon as 150 ms after a stimulus appears. The N170 component is larger when the stimulus is a face as compared to a non-face object [73].

The N200 is often linked to language control mechanisms and appears around 250-350 ms after stimulus [89, 90, 91, 92]. There are conflicting findings on the N200. Some switching studies connected the N200 to inhibitory control during bilingual speech production [89, 90, 91, 92]. Different switching costs, where switching into L1 is more costly than into L2, were attributed to the N200 [91], but this effect was also found in the other direction [89]. To investigate how performance of subjects in one language was influenced by the use of another language beforehand switching tasks were carried out, which were accompanied by a N200 component. They showed that L1 was inhibited during naming in the L2 and that this had a negative after-effect when naming the same pictures in L1 later [90]. In conflict tasks large stimulus-locked N200 were found on incongruent trials [84, 85]. But these error-response negativities can also appear on correct trials, if there is a high response conflict [86]. They are usually smaller, called correct-response negativities (CRN) and show comparable distributions on the scalp [87]. It is argued, that ERN and CRN are part of the same component that displays a comparison process that occurs before error detection, or even an emotional response to an error [88].

The N400 component was found in 1978 when an extended oddball experiment to study language was conducted [74]. They used simple sentences, either with meaningful or with semantically anomalous endings. The anomalous words elicited a large centro-parietal negativity peaking around 400 ms. It seems to reflect costs of lexical activation (ease of effort; easier accessed memory has a reduced N400) or semantic integration, such as in words terminating low versus high contextually constrained sentences [74, 75]. Other findings show, that when word-pairs are used in a semantic priming paradigm, the N400 was reduced when the second word was related to the prime [76]. Also concrete words elicit a larger N400 rather than abstract words and to words with higher density orthographic neighborhoods [77].

Violation paradigms are a common way of studying morpho-syntactic processes. These violations are thought to decrease or increase the workload of the brain's systems responsible and should show a differential ERP. Such as the early, usually left lateralized anterior negativity (LAN) between 100 and 500 ms, that is linked to automatic first pass parsing [97].

The late centro-parietal positive component (LPC) between 500 and 1000 ms, also called P600 or syntactic positive shift (SPS) [78] reflects controlled attempts to reanalyze and fix anomalies and is generally a marker for structural processing. The LPC (or 'old/new' effect) indicates recollection of information in long-term memory [79]. Words primed by repetition and words encoded deeply versus shallowly display a larger LPC during the recollection of a word [79, 80, 81]. Imageable words produce a larger LPC effect, compared to non-imageable words that are accessed from long-term memory [82]. In this way, an iconic gesture may create a deeper and stronger memory trace, because it creates a more imagistic memory for the meaning than speech alone.

1.3.3. ERD/S

Besides ERPs events also cause changes in the ongoing EEG as event-related desynchronization (ERD) or event-related synchronization (ERS). They are highly frequency-band specific, not phase locked to the event and can be detected by frequency analysis. ERDs are short, localized decreases of amplitude of rhythmic activity that occur for example when occipital alpha rhythm is blocked after visual stimulation or the central mu rhythm is blocked with active or passive movement [123]. Furthermore a higher magnitude of ERD can be found in task with greater complexity or attentional demands [124, 125]. Mu rhythms that are enhanced during visual stimulation [128], the beta rebound after limb movement [129] or the gamma activity that arises during visual processing [130] are examples for ERS and represent increase in amplitude of rhythmic activity [126]. The amplitude of fluctuations decreases with increasing frequency, because the amplitudes are proportional to the number of synchronously active neural elements [127].

The power within frequency bands is compared to the power of the baseline period before the event. Computation of ERD time course of band power values requires several steps:

1. all event-related trials are bandpass filtered
2. amplitude samples are squared to gain power samples
3. power samples are averaged across all trials
4. samples are averaged over time, for smoothing and reduction of variability in data.

With this process a phase-locked power increase of an ERP can mask the non-phase-locked power decrease (ERD). To differentiate between phase-locked and not phase-locked power changes steps 2 and 3 are replaced by the calculation of the point-to-point inter-trial variance [131]. Subject-specific frequency bands are found by comparison of short-time power spectra for reference period and activity period.

The difference curve between them can then be used to find the significant frequency components [132] that are depicted as percentages with equation 1.

$$\mathrm{ERD\%} = \frac{(\mathrm{A} - \mathrm{R})}{\mathrm{R}} * 100 \qquad \textbf{Equation 1}$$

A is the power of a frequency band following an event and R is the power of the baseline [133]. To express absolute band power as a percentage the power of the reference interval is defined as 100% and a decrease in power represents an ERD while an increase represents an ERS [126].

By distinguishing between different frequency bands of the extended alpha band a Lower alpha desynchronization (roughly 8 to 10 Hz) and a Upper alpha desynchronization (roughly 10 to 12 Hz) can be observed. The former is widespread across the scalp and a response to many tasks, while the latter is topographically restricted and related to task specific aspects [134, 135]. Prior to auditory feedback blocking of alpha band rhythms can be observed, indicating anticipation of visual or auditory feedback stimulus [136]. A significant ERS is elicited in both alpha frequency bands by auditory memory tests [137]. Low-amplitude desynchronization of EEG is caused by an increase in cellular excitability in thalamocortical systems [138]. If more cell assemblies are involved in information processing or a bigger neural network is active, larger and more spatial distributed ERDs are the result. This can be the case for increased task complexity or efficiency in performance [124, 125, 139, 140]. The upper alpha frequency range is attributed to sensory–semantic memory processes of a long-term memory system, and the lower alpha band to attentional processes [134].

Increased language processing demands cause an increase in working memory systems load. This is reflected in a selective increase of theta activity, while an increase in semantic processing is reflected in a selective decrease of upper alpha activity. Increased demands for language processing caused increased theta activity, while increased semantic processing caused decreased upper alpha activity [141].

1.4. Goal of the study

The objective of the thesis was to determine the influence of gestures on memory performance of a language task (novel abstract word learning) by means of augmented reality and EEG. For this purpose a study was designed, where subjects learned with or without gestures in multiple sessions with the virtual agent as a teacher. EEG paradigms were devised with a focus on translation in both directions.

The gesture group had to learn the corresponding movements in addition to the words and was expected to have a disadvantage in the first days of testing. But as in other studies with abstract words we expected that the gesture group will perform better than the group without gestures in the long term [22]. An incremental learning performance was expected for all subjects over the course of the study and a drop in memory performance after waiting period. Furthermore the task free recall of German words and translation from Vimmi to German were expected to yield the best results since they involve the native tongue the most and can therefore be considered the easier tasks.

This study is a further step of using augmented reality and virtual agents in a learning environment.

2. Methods

2.1. Participants

30 healthy participants (12 female and 18 male) took part in this study. They all had normal, or corrected to normal vision, were native German speakers and their ages ranged from 21 to 30 years (M = 26.1, SD = 2.88). All of them received a payment for attending the study. 28 participants were in the process of getting an education at University level and two had already finished their education in the previous year.

Groups were randomly assigned for every subject and gender, as they entered the study, resulting in 7 females and 8 males in group A (15 overall), 5 females and 10 males in group B (15 overall).

2.2. Procedure

To examine the influence of gestures during learning of foreign words, subjects tried to memorize different artificial words with or without gestures. In both groups they memorized the words with the aid of the virtual agent VARA (figure 5) in an augmented reality environment in multiple sessions.

A between-subject design was selected and the participants of the study were divided into two groups: Group A: With gestures; Group B: Without gestures. Subjects had to memorize the collection of words on three days, each followed by a rest day. After each session they completed a Recall test to track their progress. Following a final Rest period of one or two days the process of translation was examined using EEG.

To determine the sustained yield of learned and retained words, a final Recall test was carried out 30 days after the EEG Test.

As indicated by Figure 4 the whole study took place over a time span of 7 days plus one day for the 30 Day Recall.

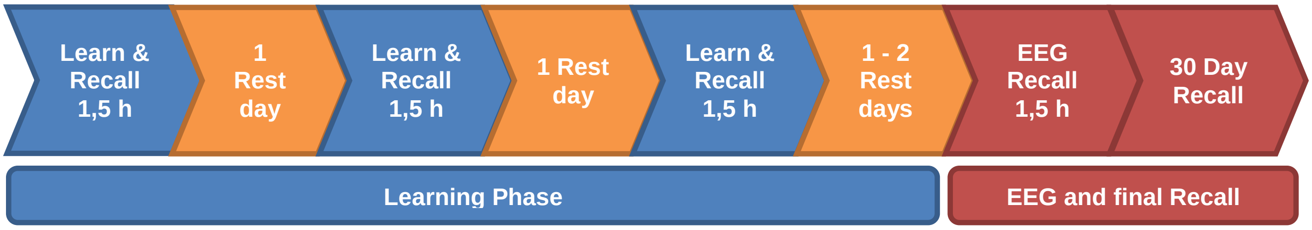


Figure 4: Course of events for the study

Figure 5: Screenshot of VARA from the software Abstract Reality

2.2.1. Learning Phase

In every Learning Phase each subject was given the task of memorizing 27 Vimmi-words of an artificial corpus created for experimental purposes [83, 22]. In detail, they consisted of 17 adverbs, six subjunctions, two conjunctions and two particles. The participants learned with the help of the virtual trainer called VARA in an augmented reality environment. At the beginning, VR-Goggles were mounted and the software “Abstract Reality” displayed VARA on a wall two meters in front of them.

Group A had to stand, because the gestures involved movements of the upper and lower body. Participants of Group B were seated in the same distance and told to remain still.

Depending on the settings VARA would cite nine Vimmi words repeatedly and in varying order for 15 minutes according to one of the three starting blocks (see Table 1). While for group A (with gestures) VARA performed explicit gestures, which have been imitated by the participants, group B received no movement instructions and VARA stands still.

Additionally the Vimmi word and its German translation were displayed in the upper left corner of the visible area. Every word was presented 12 times. Between each word block there was a 5 minute break.

Table 1 Overview of learning blocks, with all German-Vimmi pairs

Block 1		Block 2		Block 3	
Vimmi	**German**	**Vimmi**	**German**	**Vimmi**	**German**
baku	nicht	fegla	noch	lapo	wo
muladi	erst	nibesa	weil	gubame	bevor
liwe	falls	dero	sobald	ifra	schon
elebo	sogar	sigule	danach	atesi	mit
nagri	seitdem	itru	trotzdem	doza	gerade
serawo	aber	mewima	oder	wiboda	nur
koga	bis	ziso	wann	utike	entweder
dafipo	ob	pamagu	nie	kune	fast
teni	ziemlich	egi	auch	bilu	zumal

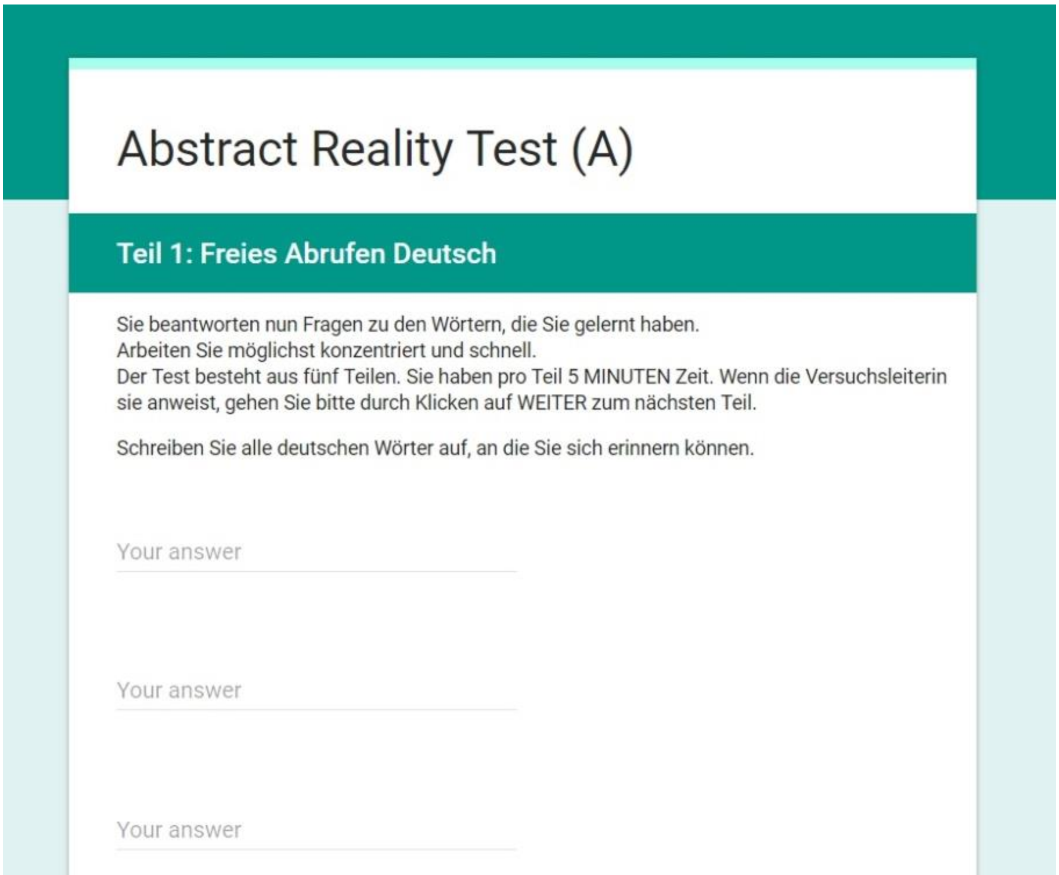
2.2.2. Recall Test

After every learning unit and another five minute break all subjects were told to perform the Recall Test on a standard PC, to determine their learning progress. It was divided into five sub phases which had to be completed within five minutes or less:

1. Free recall of learned German words
2. Free recall of learned Vimmi-words
3. Free recall of learned German-Vimmi pairs
4. Translation of learned words from German to Vimmi
5. Translation of learned words from Vimmi to German

In the free recall tests, participants were presented with empty text fields, with the instruction to fill in as many words or word-pairs as they were able to remember. In the translation tests subjects were provided with all 27 words, which they had to translate in both directions.

Figure 6 gives an example of the Google Forms for the Recall Phase



Abstract Reality Test (A)

Teil 1: Freies Abrufen Deutsch

Sie beantworten nun Fragen zu den Wörtern, die Sie gelernt haben.
Arbeiten Sie möglichst konzentriert und schnell.
Der Test besteht aus fünf Teilen. Sie haben pro Teil 5 MINUTEN Zeit. Wenn die Versuchsleiterin sie anweist, gehen Sie bitte durch Klicken auf WEITER zum nächsten Teil.

Schreiben Sie alle deutschen Wörter auf, an die Sie sich erinnern können.

Your answer

Your answer

Your answer

Figure 6 Screenshot of the first Recall Test made with GoogleForm

2.2.3. 30 day „Final Recall“

30 days after the EEG-Recall all subjects were sent a link to a Recall test via email. It was identical to the ones they had already completed after every Learning Phase. This was done for a final determination of their long-term memory performance.

2.2.4. EEG Measurement (Experimental Paradigms)

EEG measurements took place after one or two rest days following the learning phase (see figure 4) and were equal for all subjects. WaveGuard^{TM} electrode caps (in sizes S, M or L) from ant neuro with 64 active electrodes, arranged in the international 10-20 system were used. Ground was located in the middle of the forehead and the reference electrode was CPz. The cap was attached to the eego^{TM} amplifier from ant neuro (EE-224; eemagine Medical Imaging Solutions GmbH) which was in turn connected to a tablet with the resistance measurement software eego^{TM} (LE-200; Version 1.5.6 from eemagine Medical Imaging Solutions GmbH) to check the electrode resistance. Electrolyte paste (OneStep ClearGel) was applied to decrease the electrode resistance under a value of 5 kOhm. The paste was easily removed after the trials by simple hair washing.

Afterwards the cap and the amplifier were connected to an IBM-Laptop to record the EEG, as well as displaying the live EEG continuously to check for artifacts or errors during recording. Another IBM-Laptop displayed the paradigm programs pyEEG. The four paradigm GUIs were written in PyCharm (Version 2016.3.2) using Python 2.7.13 and the Python API for the lab streaming layer PYLSL version 1.10. After the paradigm was finished, pyEEG created .txt files with the subjects' answers, which were needed for further analysis. Recorded EEG and Triggers provided by the paradigms were sent to the Post Receiver and were stored in the .xdf Fileformat (Extensible Data Format).

The subjects were instructed not to roll with their eyes, not to blink and not to tense their facial-, or other muscles to prevent disturbing signals (artifacts). Those artifacts were demonstrated to the subjects by letting them perform those unwanted movements and showing them the visible changes in a live EEG stream. Furthermore they were told that there would be pause periods in every paradigm in which they could relax. The EEG measurements took about 1.5 hours.

Before each test subjects performed a rehearsal of the paradigm to familiarize themselves with the trial cycle. The following EEG-Recall tests were conducted:

1. Free recall of learned German-Vimmi pairs
2. Translation of learned words from German to Vimmi
3. Translation of learned words from Vimmi to German
4. Identification of correct German – Vimmi pairs

2.2.5. Free recall of learned German-Vimmi pairs

On the start screen the instruction to press “ENTER” appeared. Afterwards a fixation cross was shown for two seconds and subjects were told previously to focus on it. After that the reflection period started, in which the participants were prompted with the request to think of a German-Vimmi pair.

By pressing “ENTER” again, the subjects ended the reflection period and the program moved forward to the entry screen, where the participants were asked to input their word pair into a text field. This input period also served as the pause period for this paradigm. When the subjects felt ready they confirmed their entry with “ENTER” and the fixation cross was prompted again, starting the process anew.

Participants were told to continue entering word pairs, until they were no longer able to think of any. At this point, they were able to conclude the experiment in the reflection phase by pressing “ESC”. The timing of what was displayed is exhibited in Figure 7.

Besides “ENTER” for continuation, “ESC” for concluding the test and the entries into the text field, no other inputs were possible.

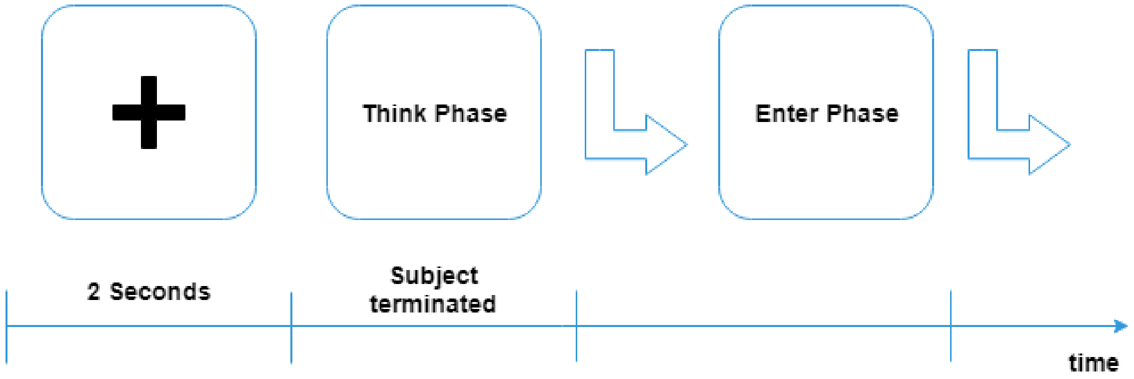


Figure 7: Timeline of EEG paradigm Free recall of learned German-Vimmi pairs

2.2.6. Translation of learned words from German to Vimmi

The program was similarly built to the first one and showed a start screen which asked the subject to start the paradigm by pressing “ENTER. Afterwards a fixation cross was shown for two seconds. Next a German word was displayed for five seconds, after which the program moved forward to the entry screen that is equal to the one in the previous paradigm. The period in which the word is displayed could be cut short if the participant pressed “ENTER” in case they were able to translate the word faster.

In the entry screen the subjects were asked to input their translated word into a text field. This input period also served as the pause period for this paradigm. When the subjects felt ready they confirmed their entry with “ENTER” and the fixation cross was prompted again, starting the process anew. If the participants were not able to translate the word, they were told to simply leave the text field empty.

Participants were presented each of the 27 words and were prompted the final screen, informing them that the test was concluded afterwards. The timing of what is displayed is exhibited in Figure 8.

Besides “ENTER” for continuation and the entries into the text field, no other inputs were possible.

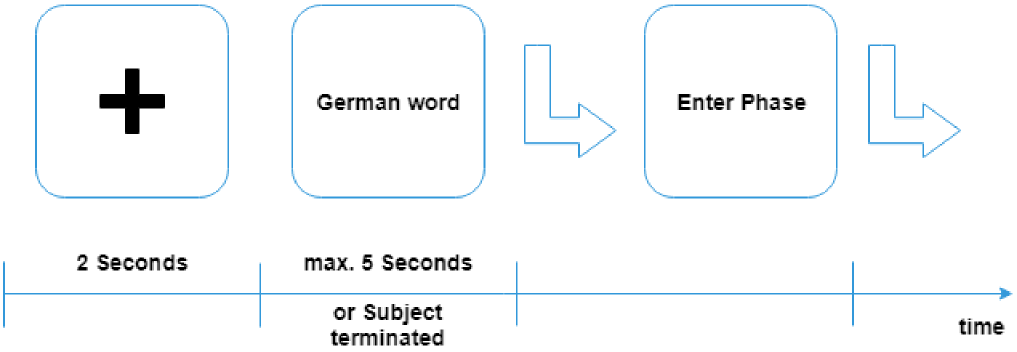


Figure 8: Timeline of EEG paradigm translation of learned words from German to Vimmi

2.2.7. Translation of learned words from Vimmi to German

The paradigm to translate from Vimmi to German is equal to the previous one, besides prompting Vimmi instead of German words.

2.2.8. Identification of correct German – Vimmi pairs

The last paradigm also follows the design of the previous ones and was started by pressing “ENTER”. A fixation cross was displayed for two seconds, followed by a German – Vimmi pair. By pressing the arrow keys (Right = TRUE; Left = FALSE) the subjects decided whether the words were paired correctly or not. This was followed by a pause screen, which served as the pause period for this paradigm. By pressing “ENTER” again, the pause was concluded and the fixation cross was prompted again, starting the process anew. Other inputs were prohibited.

Subjects were polled all 27 correct word pairs and 10 incorrect word pairs in a randomized order, after which the End Screen was displayed that concluded the test. The timing of what was displayed is exhibited in Figure 9.

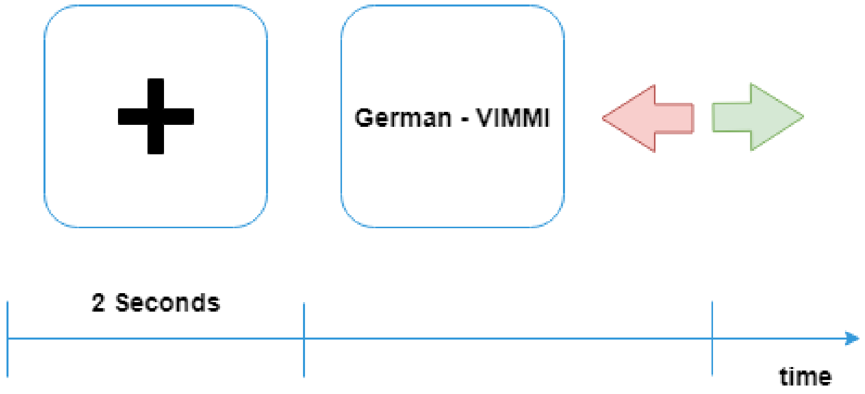


Figure 9: Timeline of EEG paradigm identification of correct Vimmi – German pairs

2.3. Data Analysis

In a first step of data analysis the durations which subjects needed to answer each single task were reviewed. This was done for all four paradigms (Figure 16 Appendix).

All preprocessing activities and data analysis of the EEG-data were completed using Matlab toolbox EEGLab. eeglab() is a Matlab GUI environment for electrophysiological data analysis incorporating the ICA/EEG toolbox. The plugin "xdfimport" v1.13 for eeglab was used to loead .xdf –EEG datafiles.

2.3.1. Preprocessing

The preprocessing of the EEG-Data was conducted by a series of steps (see Figure 10), involving manual inspection and automated processing of the data. The following descriptions were applied to all EEG-data of all paradigms.

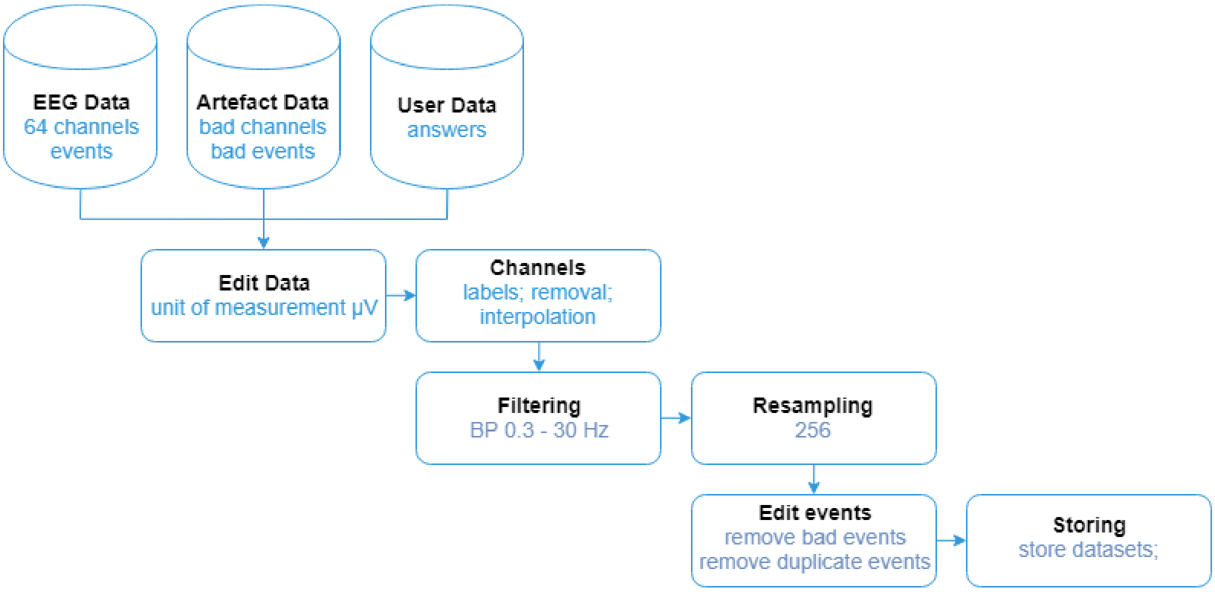


Figure 10: Flowchart of data preprocessing steps

To be able to remove trials with muscle or blink artifacts and noisy channels in later stages of preprocessing, the data was manually inspected. EEG and events were loaded with EEGLAB and a basic bandpass filter of 1 Hz to 100 Hz was applied. Then the EEG was plotted and each trial and channel was reviewed. The number of “bad events” and the name of “bad channels” were then stored in separate .txt files. Subsequently these were not used further on.

Each EEG paradigm produced a .xdf-file with the EEG data and the event triggers, as well as a .txt-file containing the answers of the subject. These and the two .txt-files containing “bad channels” and “bad events” were loaded. Further on only correct answers and valid trials were used.

The raw EEG data had to be multiplied by 10^6 to change the unit of measure from Volts to micro Volts.

Labels for each channel were loaded from an extra file, since the .xdf-files don’t contain this information. Afterwards “bad channels” and the reference channel CPz could be removed by name and interpolated using surrounding channels.

To narrow the frequency band a FIR bandpass filter (finite impulse response) of EEGLAB with cut-off frequencies $f_{c,low}$ = 0.3 Hz and $f_{c,high}$ = 30 Hz was applied for ERP analysis of the signals.

To reduce further calculation time, the EEG-data was down sampled to decrease the size of the data from 512 Hz to 256 Hz. This value was chosen, because the information we are looking for can be found in a low frequency range, therefore not altering the results. For anti-aliasing purposes resampling took place after the filter process was completed.

During recording of the EEG, the Lab streaming layer stored two streams with event information, which can be found as identical events in EEGLab. These duplicate events had to be removed, as well as false answered trials.

2.3.2. Learning Progress

To analyze the overall learning results and the progress over time, the answers of each subject of every recall test was compared to the correct answers in an automated R-Script. An absolute match was given a score of 1 and a false answer was given a score of zero. An empty field was also assigned a value of 0. Afterwards all answers with a score of zero were manually reviewed. The score was changed to a value of 0.5 if one of the following conditions was met:

- One or two missing letters
- One or two extra letters
- Switched letters or syllable
- Two of three syllables are correct

The scores were added up for each subject and each recall test, providing a numerical representation (from 0 to 27) of the corresponding learning progress.

First graphs were created for each group and paradigm containing the scores of every person for every recall test. Afterwards the median and boxplots of those values for each recall test day were calculated, so that the two groups could be compared side by side (See Appendix).

In the next step a median-split for each group was conducted, separating the high- and low-performers within a group at the median value. New median values for all four groups were calculated, as well as new boxplots. Those were again compared in graphs.

2.3.3. Grand Average ERPs analysis

After preprocessing, the EEG data was epoched and baseline corrected.

EPOCHing reveres to segmenting the whole EEG dataset into small parts around the stimulus (-2000 ms pre and 5000 ms post stimulus), so that each segment corresponds to a single trial. The stimuli were the start of the thinking-phase for paradigm 1, prompting of words for paradigm 2 and 3 and prompting of pairs for paradigm 4.

After subtraction of the mean signal value pre-stimulus (-1500 ms to -500 ms) from the signal post-stimulus for baseline correction, the datasets were saved and used to create EEGLAB-studies.

2.3.4. Statistical analysis

The groups will be summarized with the following abbreviations:

A High:	High performers of subjects learning with gestures
B High:	High performers of subjects learning without gestures
A Low:	Low performers of subjects learning with gestures
B Low:	Low performers of subjects learning without gestures

To determine the influence of gestures on memory, three 4x5 ANOVAs for repeated measurements with TIME (day 1, day 2, day 3, day 4) and TCOND (test condition: Gfree, Vfree, Pairfree, tGV, tVG) as within-subject factors were conducted.

Between-subject factor for the first ANOVA was GROUP (A High, B High, A Low, B Low). Between-subject factor for the second ANOVA was GROUP (A High and group B High) and between-subject factor for the third ANOVA was GROUP (A Low and B Low).

For every paradigm a Grand Average ERPs-Graphs were generated using the average of all subjects of one group, resulting in two ERPs curves from -100 ms pre- to 1000 ms post-stimulus, for the electrode positions P7 and CP5. Differences in ERP curves are then analyzed with the aid of Welch Two Sample t-tests for N200 and N400.

3. Results

In the first section of this chapter behavioral memory results are presented and graphically displayed. The next section shows statistical analysis of the EEG results of all participants for selected electrode positions.

3.1. Behavioral Data

The overall results of the memory tests divided by group and high/low performers are shown as median number of correct answers for each condition over the four test days. Similarly boxplots were created to gain a clearer insight into the data.

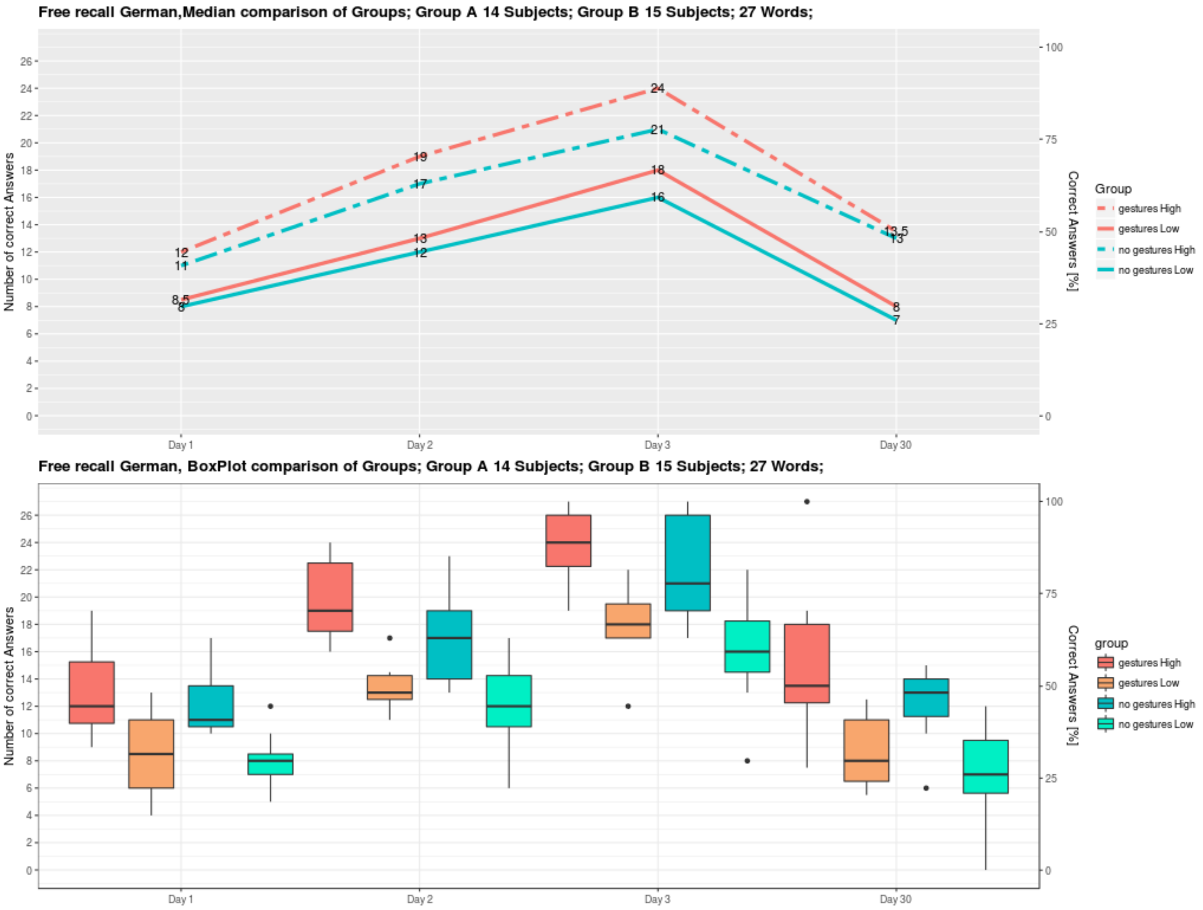
Results from subjects enacting gestures during learning phases are displayed as red lines, or light red boxplots for low performers and dashed red lines, or red boxplots for high performers. Blue and dashed blue lines, as well as light blue and blue boxplots represent low and high performer results from subjects that did not learn with gestures.

The differences between the groups that were analyzed with ANOVAs are also reviewed and are shown in 3.1.6

3.1.1. Free recall of learned German words

The results of the ANOVAs can also be seen in the graphic representations of the learning results. But A High and A Low seem to have a slight advantage over B High and B Low for recollection of German words throughout the study.

Figure 11 depicts results for Free recall of learned German words for groups A High, B High, A Low, B Low. The upper half shows median learning results over the course of the three learning days, as well as day 30 after EEG measurement. The lower half consists of boxplot of all four groups for the same points in time.



Free recall German,Median comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

Free recall German, BoxPlot comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

Figure 11: Results for Free recall of learned German words

3.1.2. Free recall of learned Vimmi-words

Median results for this test indicate an advantage for the B Low over A Low during Learning Phases. But they lose their advantage in the 30 day Recall test. Also the boxplots indicate a higher variability in learning results for B High, than A High.

Figure 12 depicts results for Free recall of learned Vimmi words for groups A High, B High, A Low, B Low. The upper half shows median learning results over the course of the three learning days, as well as day 30 after EEG measurement. The lower half consists of boxplot of all four groups for the same points in time.

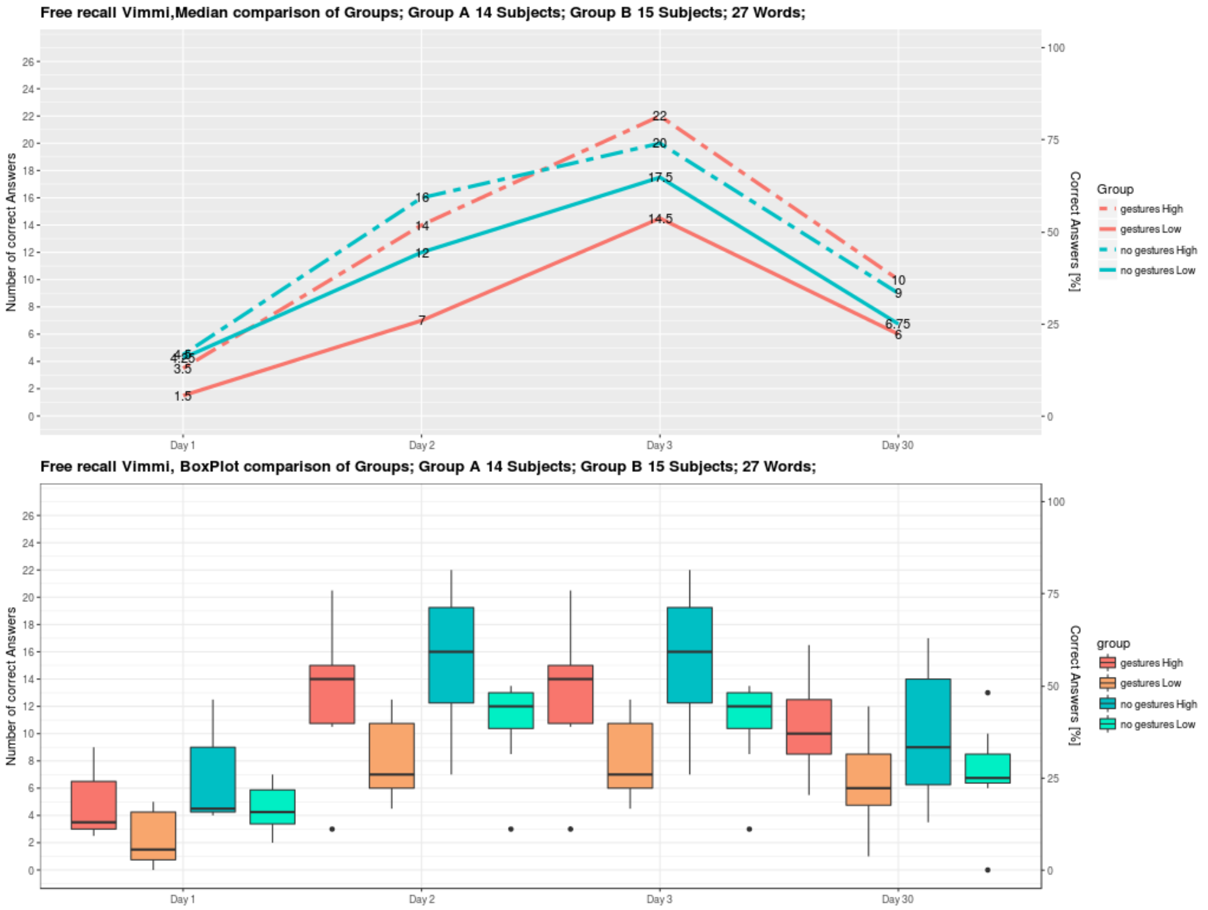
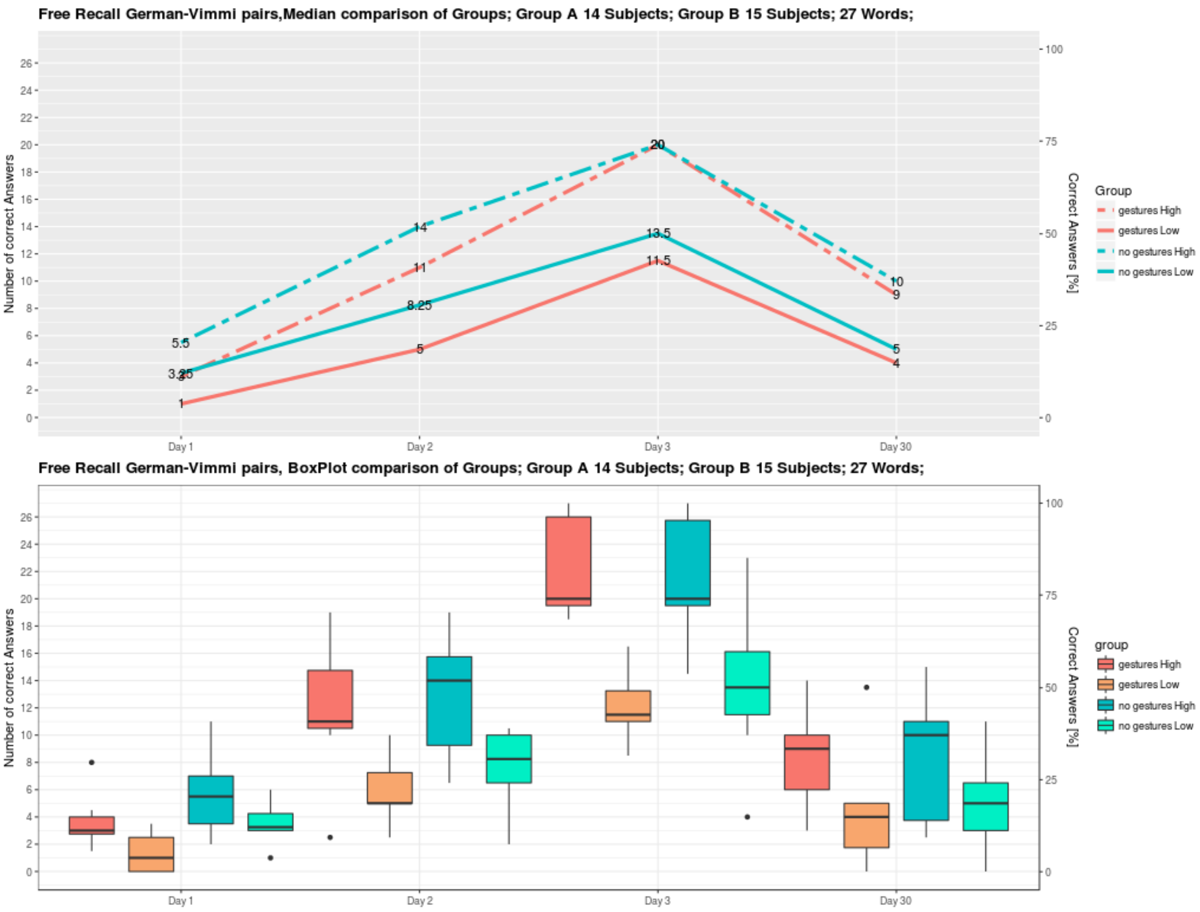


Figure 12: Results for Free recall of learned Vimmi words

3.1.3. Free recall of learned German-Vimmi pairs

As indicated by the ANOVA results no big differences can be seen between the two high performer groups, but B Low subjects have a higher median result throughout the study, than A Low. Again B High display a higher variability in results than their gesture counterparts in the boxplots.

Figure 13 depicts results for Free recall of learned German-Vimmi pairs for groups A High, B High, A Low, B Low. The upper half shows median learning results over the course of the three learning days, as well as day 30 after EEG measurement. The lower half consists of boxplot of all four groups for the same points in time.


Free Recall German-Vimmi pairs,Median comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

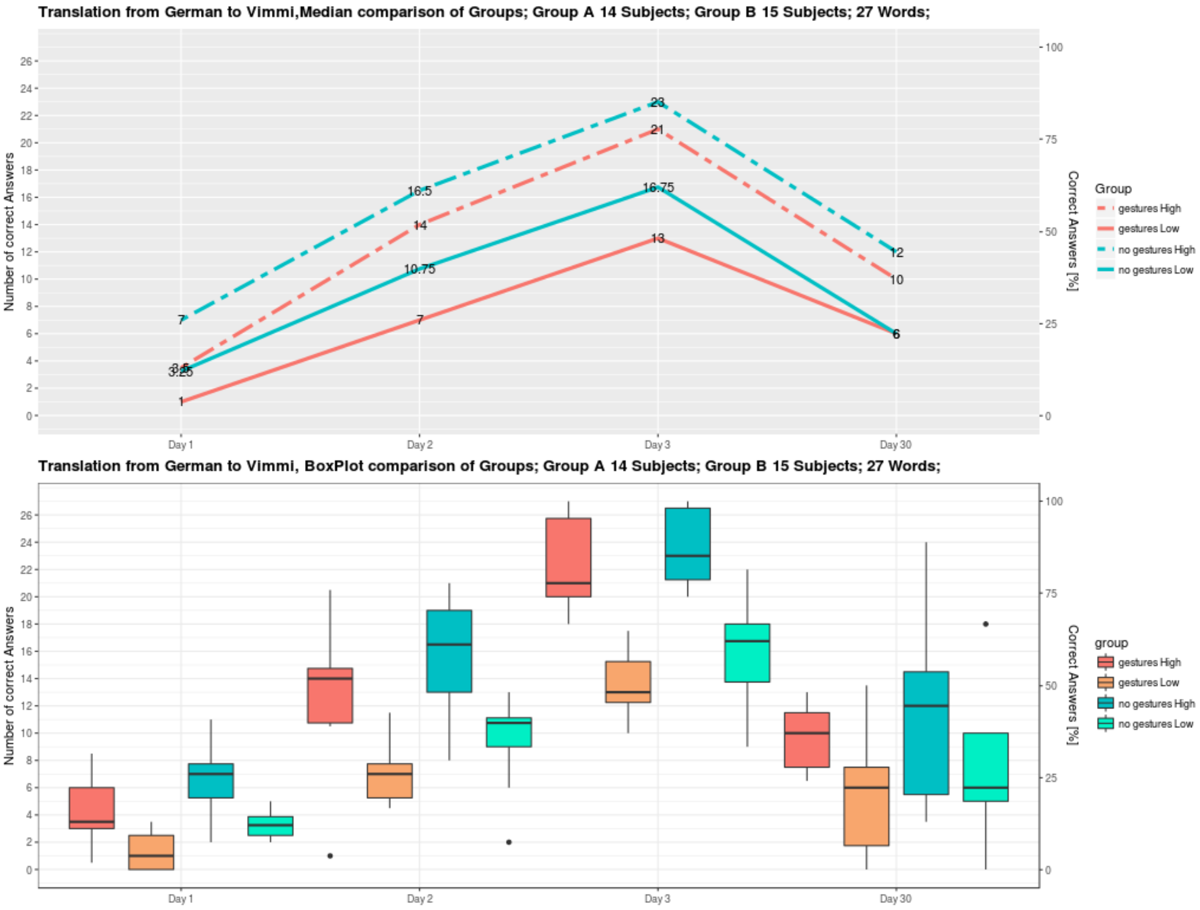
Free Recall German-Vimmi pairs, BoxPlot comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

Figure 13: Results for Free recall of learned German-Vimmi pairs

3.1.4. Translation of learned words from German to Vimmi

In the first translation task, subjects without the aid of gestures have a slight advantage, which dimishes after 30 days. Again B High display a higher variability in results than their gesture counterparts in the boxplots.

Figure 14 depicts results for the task translation of learned words from German to Vimmi for groups A High, B High, A Low, B Low. The upper half shows median learning results over the course of the three learning days, as well as day 30 after EEG measurement. The lower half consists of boxplot of all four groups for the same points in time.


Translation from German to Vimmi,Median comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

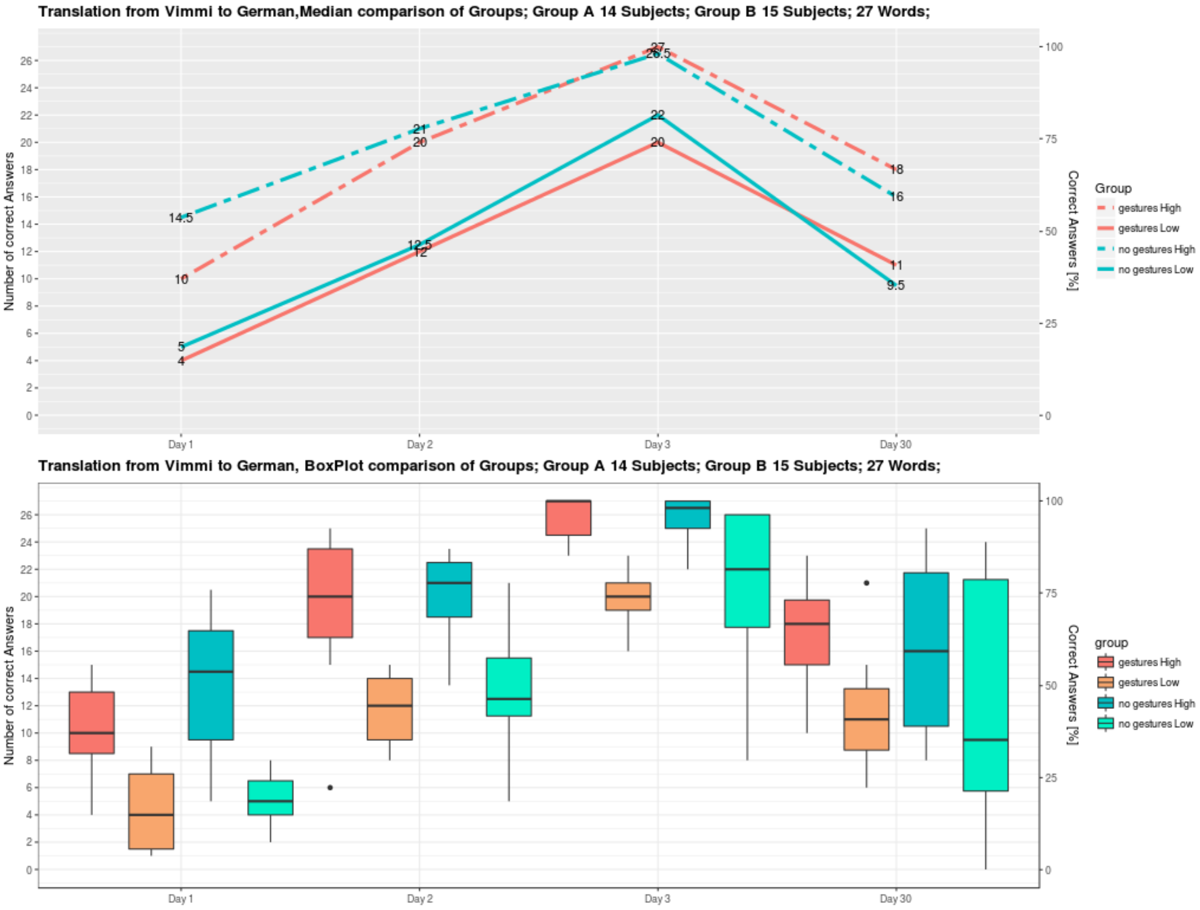
Translation from German to Vimmi, BoxPlot comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

Figure 14: Results for task translation of learned words from German to Vimmi

3.1.5. Translation of learned words from Vimmi to German

Translation from Vimmi to German produced the best learning results with both high performer groups achieving a median result of almost the maximum score. Results between high and between low performers were very similar throughout the study, with the gesture groups retaining more information after 30 days. The no gesture groups display a higher variability in results, especially at the 30 day mark.

Figure 15 depicts results for the task translation of learned words from Vimmi to German for groups A High, B High, A Low, B Low. The upper half shows median learning results over the course of the three learning days, as well as day 30 after EEG measurement. The lower half consists of boxplot of all four groups for the same points in time.



Translation from Vimmi to German,Median comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

Translation from Vimmi to German, BoxPlot comparison of Groups; Group A 14 Subjects; Group B 15 Subjects; 27 Words;

Figure 15: Results for task translation of learned words from Vimmi to German

3.1.6. Results of the ANOVAs for repeated measurements

A) Repeated measures ANOVA High and Low performers

There was a significant effect of within-subject factors TCOND (F(2.089, 52.223) = 60.665, p<.05), as well as TIME (F(1.651,41.267) = 95.512, p <.05) for words remembered. The between-subject factor GROUP showed also a significant effect (F(3,25) = 7.383, p < .05). There was also a significant interaction between TCOND and TIME (F(5.72,142.998) = 12.267, p < .05), as well as between TCOND, TIME and GROUP (F (17.16,142.998) = 2.545, p < .05)

B) Repeated measures ANOVA High performers

There were again significant effects for within-subject factors TCOND (F(2.253, 27.041) = 26.3, p<.05) and TIME (F(1.599, 19.192) = 47.762, p <.05), as well as their interaction (F(4.994, 59.927) = 10.54, p <.05). But there was no longer a significant effect for GROUP or any interaction with GROUP.

C) Repeated measures ANOVA Low performers

Similarly to high performers, there were again significant effects for within-subject factors TCOND (F(1.594, 20.72) = 35.945, p<.05) and TIME (F(1.67, 21.715) = 50.169, p <.05), as well as their interaction (F(3.785, 49.21) = 7.228, p <.05) was found. But there was no longer a significant effect for GROUP or any interaction with GROUP.

As expected, the repeated measures ANOVAs revealed, that there was a clear difference in memory performance for Factor TIME, indicating the learning progression. The same can be seen for the different testing conditions (TCOND), due to varying difficulty of tests. After the median split no difference between A High vs. B High and A Low vs. B Low was detectable.

3.2. EEG Data

Here selected electrode positions are being presented according to expected N200 and N400 locations (CP5 and P7), for the translation tasks and for each group (A red line; B green line). The black vertical line at 0 ms represents the onset of the stimulus (word that is to be translated). For the translation tasks, paired sample t-tests for all participants were conducted in order to determine whether the minimum values of the N200 and N400 were significantly different. This was done using the MATLAB® function t-test assuming unknown and unequal variance, since the number of subjects used was unequal. The minimum values in the interval of 100 to 300 ms after the stimulus were used for N200 and from 300 to 500 ms after the stimulus were used for N400 as inputs for the t-tests.

3.2.1. ERPs during free recall of German-Vimmi pairs

The free recall was conducted in order to prepare the subjects for the tasks ahead to familiarize them with the following paradigms and also to keep in line with the previous Recall Tests. The input times, subjects needed to remember Vimmi-German pairs, were analyzed and showed high variability (Figure 16). Therefore accurate ERP analysis was not possible.

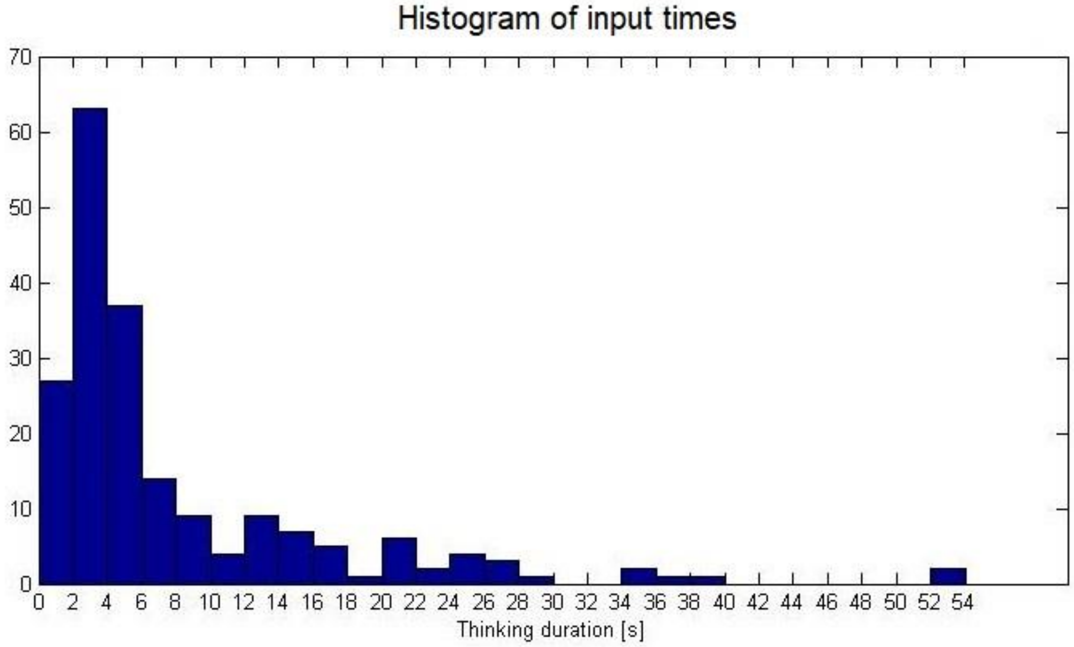


Figure 16 Histogram of input times, subjects needed to remember German-Vimmi pairs

3.2.2. ERPs during translation from German to Vimmi

The aim of this paradigm was to look for differences in ERP curves between the gestures and no gesture group for correct answers on a translation task from German to Vimmi. Figure 17 shows grand average ERPs of all participants at electrode positions CP5 and P7 from 100 ms pre-stimulus to 1000 ms post-stimulus separated for groups A and B. A Welch Two Sample t-test was conducted to compare whether there was a difference between subjects that learned with or without gestures for N200 and N400 components.

Table 2: T-test results for ERPs during translation from German to Vimmi

ERP component	Electrode	Group	M	SD	t	df	p
N200	CP5	A B	-5.2119 7.4297	4.3276 1.8044	1.6479	14.472	0.1209
N200	P7	A B	-8.9792 -9.6376	7.1176 4.447	0.2747	18.191	0.7866
N400	CP5	A B	-5.4145 -7.0777	2.5733 2.8612	1.5301	22.988	0.1396
N400	P7	A B	-6.1111 -8.4396	4.4282 3.4506	1.4581	20.798	0.1597

The ERP curves for electrodes P7 and CP5 showed N200 and N400 components for both groups, but did not reveal a significant difference (table 2). The N200 component of position P7 was more distinct than its N400 component.

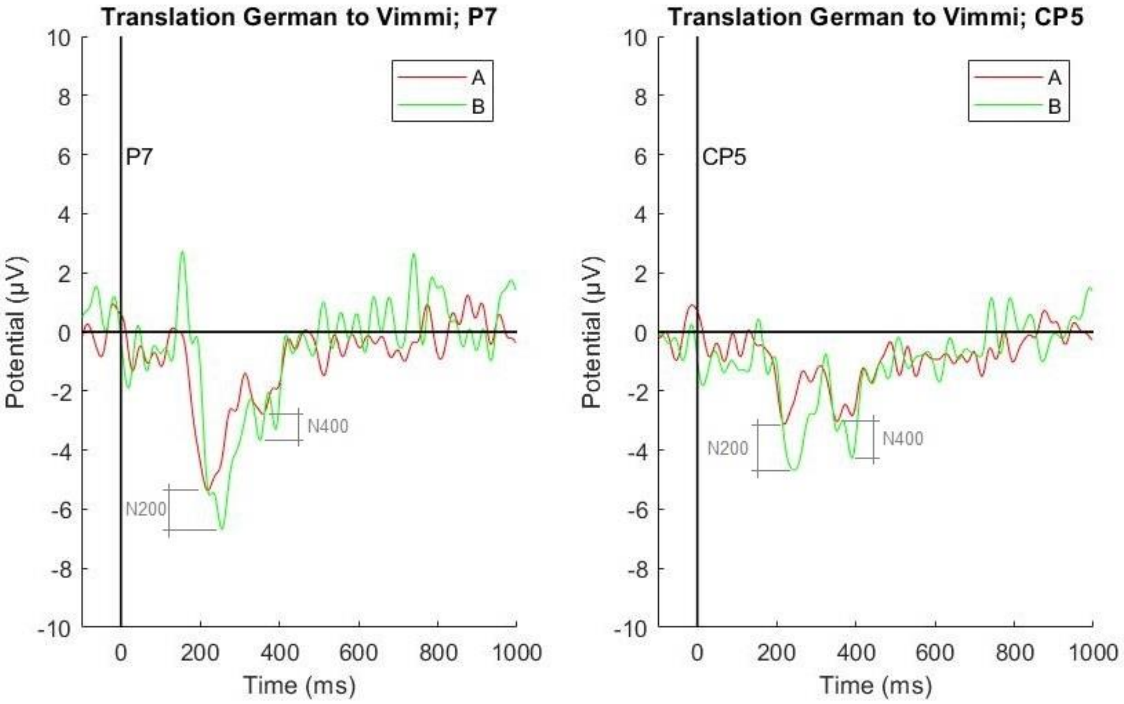


Figure 17 Grand averages of EEG signals for translation task from German to Vimmi at electrodeposition CP5 and P7 for groups A and B

3.2.3. ERPs during translation from Vimmi to German

The aim of this paradigm was to look for differences in ERP curves between the gestures and no gesture group for correct answers on a translation task from Vimmi to German. Figure 18 shows grand average ERPs of all participants at electrode positions CP5 and P7 from 100 ms pre-stimulus to 1000 ms post-stimulus separated for groups A and B. A Welch Two Sample t-test was conducted to compare whether there was a difference between subjects that learned with or without gestures for N200 and N400 ERP component.

Table 3: T-test results for ERPs during translation from Vimmi to German

ERP component	Electrode	Group	M	SD	t	df	p
N200	CP5	A B	-4.8353 -7.4731	3.4754 3.3352	1.897	21.963	0.0711
N200	P7	A B	-9.1277 -10.1176	6.6431 6.0675	0.3811	21.822	0.7068
N400	CP5	A B	-4.5457 -3.7765	2.1616 7.669	-0.3344	12.737	0.7435
N400	P7	A B	-6.0114 -7.1831	4.9336 6.3389	0.5053	20.749	0.6187

The ERP curves for electrodes P7 and CP5 showed N200 and N400 components for both groups, but did not reveal a significant difference (table 3). And again the N200 component of position P7 was more distinct than its N400 component.

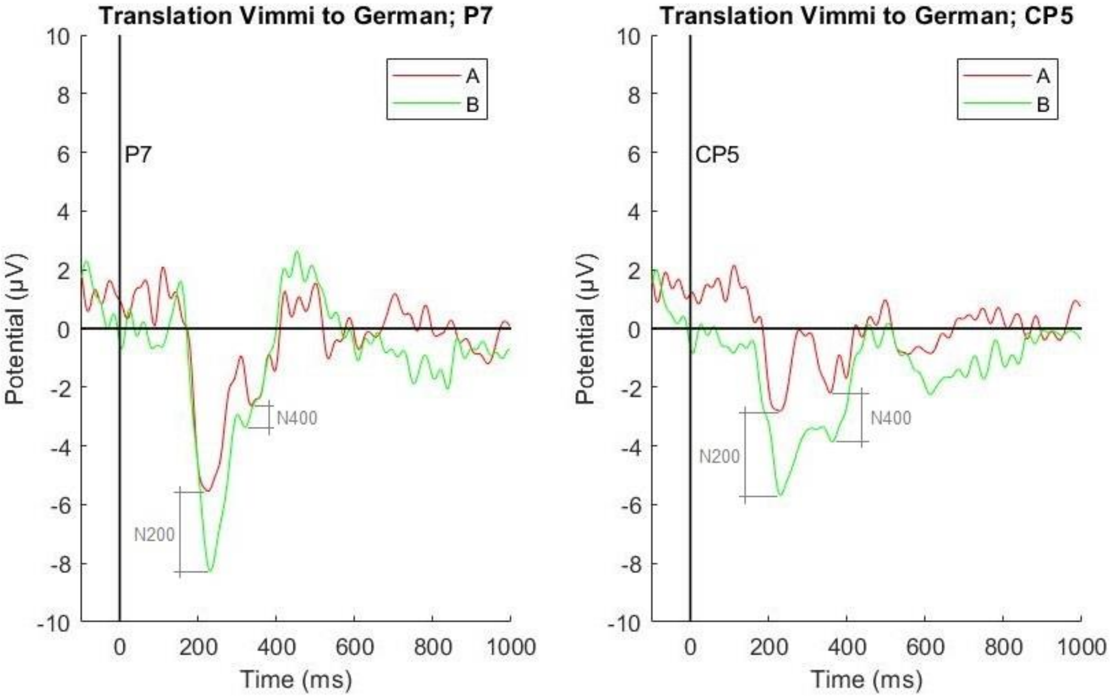


Figure 18: Grand averages of EEG signals for translation task from Vimmi to German at electrodeposition CP5 and P7 for groups A and B

3.2.4. Identification of correct German – Vimmi pairs

The aim of this paradigm was to look for differences in ERP curves between the correctly identified and falsely identified German-Vimmi pairs for both groups. Figure 19 shows grand average ERPs of all participants of groups A at electrode positions CP5 and P7 from 100 ms pre-stimulus to 1000 ms post-stimulus separated by correct and falsely identified pairs. A Welch Two Sample t-test was conducted to compare whether there was a difference for N200 and N400 ERP component between correctly identified pairs (ci) and falsely identified pairs (fi) for subjects that learned with gestures.

Table 4: T-test results for ERPs during identification of correct pairs for group A (* p < 0.05)

ERP component	Electrode	Group A	M	SD	t	df	p
N200	CP5	ci fi	-3.7994 -9.1832	1.9021 5.7059	2.879	10.346	0.0159*
N200	P7	ci fi	-6.9197 -11.7666	4.2444 5.1879	2.4567	16.688	0.0253*
N400	CP5	ci fi	-2.6743 -9.4148	1.6832 4.3154	4.7064	10.846	0.0006*
N400	P7	ci fi	-4.9359 -11.3119	2.2589 5.7046	3.3631	10.903	0.0064*

The ERP curves for electrodes P7 and CP5 showed N200 and N400 components for both cases. They were significantly more negative for falsely identified pairs (table 4).

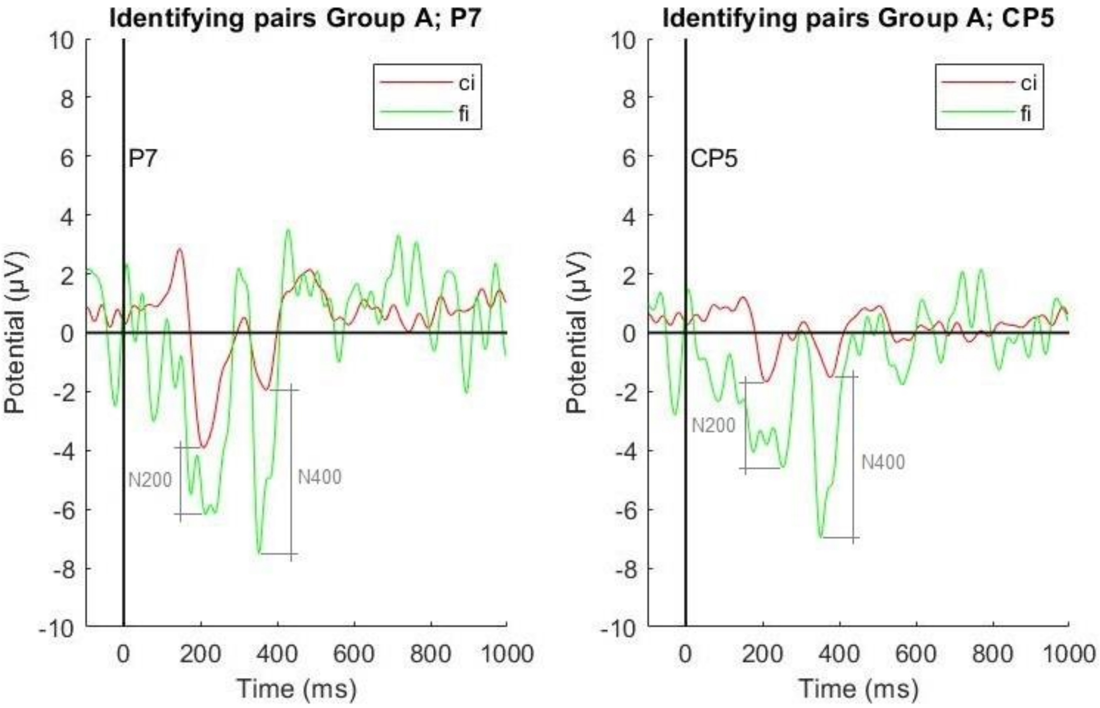


Figure 19 Grand averages of EEG signals for Identification task for Vimmi-German pairs at electrodeposition CP5 and P7 for Group A.

3. Results

Figure 20 shows grand average ERPs of all participants of groups B at electrode positions CP5 and P7 from 100 ms pre-stimulus to 1000 ms post-stimulus separated by correct and falsely identified pairs. A Welch Two Sample t-test was conducted to compare whether there was a difference for N200 and N400 ERP component between correctly identified pairs (ci) and falsely identified pairs (fi) for subjects that learned without gestures.

Table 5: T-test results for ERPs during identification of correct pairs for group B (* p < 0.05)

ERP component	Electrode	Group B	M	SD	t	df	p
N200	CP5	ci fi	-3.2898 -9.6632	1.9410 3.5480	5.3607	14.634	8.6e-05*
N200	P7	ci fi	-5.8293 -13.2477	3.3219 5.3313	4.0397	15.894	0.0009*
N400	CP5	ci fi	-2.6425 -9.0050	3.0513 3.3573	4.8947	20.538	8.1e-05*
N400	P7	ci fi	-5.0530 -11.8032	3.3215 4.9813	3.869	16.645	0.0013*

The ERP curves for electrodes P7 and CP5 showed N200 and N400 components for both cases. They were significantly more negative for falsely identified pairs (table 5).

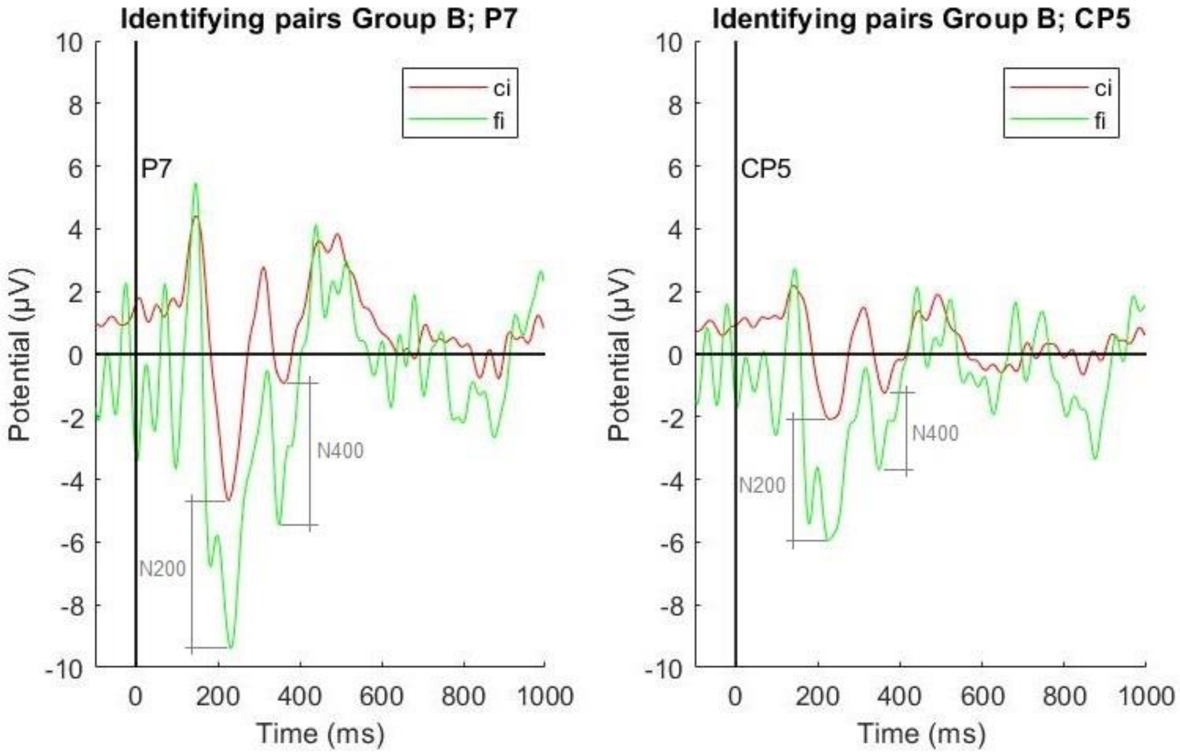


Figure 20 Grand averages of EEG signals for Identification task for Vimmi-German pairs at electrodeposition CP5 and P7 for Group B.

4. Discussion

The objective of the thesis was to show the influence of gestures on the ability to learn abstract novel words. Furthermore ERP analysis (N200, N400) was used to find differences during translation for the two groups. Since in addition to a German and a Vimmi word subjects of group A had to learn gestures, it was expected that they would have a disadvantage in the beginning, but that they would have significantly superior memory results overall. An incremental learning performance was expected for all subjects over the course of the study resulting in the best results on day three and a drop in memory performance after the 30 day waiting period. Furthermore the task free recall of German words and translation from Vimmi to German were expected to yield the best results since they involve the native tongue the most. Learning results for the five recall tests were compared visually as well as statistically. A difference in ERPs between groups and correct or incorrect answers was also predicted. Grand average ERPs for the groups were compared visually as well as with t-tests.

What follows is a summarization and interpretation of the obtained results that were achieved with the described methods, as well as improvements for future studies of this kind. The groups will be summarized with the following abbreviations:

A High:	High performers of subjects learning with gestures
B High:	High performers of subjects learning without gestures
A Low:	Low performers of subjects learning with gestures
B Low:	Low performers of subjects learning without gestures

Observable in all recall and translation tasks is the learning curve of all groups. They increased their memory performance after each learning session and it was decreased after the 30 day waiting period with no intermediate learning phases or repetitions.

Obtained memory performance for free recall of learned German words shows a very similar starting point for high performers and low performers of both groups. Throughout the testing, both A high and A low, score higher than their counterparts Similar results for free recall of German words were obtained by Macedonia and Knösche [22], where German items encoded through enactment scored higher on daily tests over a period of 6 days. Recall of German words showed the highest number of correct answers for all 3 free recalls.

For free recall of learned Vimmi-words all groups start out closely together, with group A low having the lowest memory performance. On the second recall test day both no gesture groups show advantages over the others. Group A high surpasses their counterpart on day 3 and the learning curve for group A low is steeper than that of B low. This is in accordance with other studies [22], where the superiority of enacted items only manifested after 3 days, with higher scores for enacted items after 6 days. A higher variability in memory performance can be seen in group B high throughout the testing.

Similar to free recall of Vimmi words, the free recall of German-Vimmi pairs showed lower scores for gesture groups in the beginning, but they can close this gap by day 3. Matched free recall of pairs in [22] resulted in a similar outcome, where enacted items were better recalled on day 3 till the end. As in the previous recall test, the variability of group B high is higher compared to group A high. Free recall of pairs showed the lowest number of correct answers in general for all groups.

In the first translation task from German to Vimmi results for day 1 show both groups without gestures scored higher than their gesture counterparts and results for group B low were almost equal to group A high. Group A high displays a steeper learning curve, but they were not able to catch up to group B high as they did in the free recall of pairs. This differs from [22], where cued recall from German into Vimmi showed significant advantage of enacted items from day three onwards. On day 30 the advantage group B low had over group A low disappeared and their median scores were equal. Group B high scored highest across all test days, but correct answers of its members continued to show higher variability than group A high.

The best scores were achieved for translation from German to Vimmi, with both A and B high scoring almost the full 27 translations for median results. Both low performer groups were very similar throughout the recalls, but loss of memory performance was lower for group A low after 30 days. Similarly group A high starts out worse than group B high, but catches up on day 2, surpasses them on day 3 and they retain a better memory performance on day 30. These results are in accordance to cued recall from Vimmi to German [22], with higher scores for enacted items after day 3. Yet again group A high, but also group A low, especially on days 3 and 30, displayed higher variability than their counterparts of groups B high and B low.

The high variability in correct answers for group A high across all tests indicates a more constant memory performance for subject learning with gestures. Overall a significant difference between the four groups was found, which can be attributed to the median split, but no significant difference between groups A high and B high or A low and B low was observed. This indicates no benefit or handicap for the use of gestures for three learning days. Previous studies showed a positive effect of gestures on abstract language learning [22, 95]. One of the reasons is that the superiority of enactment over audiovisual learning start with day 3 of learning and grows more prominent till day 6 [22].

It can be argued, that abstract words can only be accompanied by gestures that are seen as incongruent, meaningless, or arbitrary in connection to the word. Multiple studies showed reduced learning results if drawing of outlines [24], meaningless gestures [21], or incongruent gestures [20] were used, instead of iconic gestures. Additionally effects like the multimodal sensorimotor experience are stronger for action words [6, 41]. These reasons might have played a part in the results obtained in this study and why abstract words accompanied by gestures need longer training to show improvement over no gesture use.

While the N200 and N400 components can be visually observed, they are not significantly different between gesture and no gesture groups.

Translation from German to Vimmi shows a stronger N200 than N400 component for both groups for electrode P7, while they are of similar sizes for position CP5. However for both components and electrode position, ERPs for the no gesture groups are slightly higher.

Similarly translation from Vimmi to German shows a stronger N200 than N400 for P7 and about equal size for N200 and N400 for CP5. Again amplitudes for the no gesture groups are higher in both components and electrode positions.

This may be due to the fact that only correct answers could be evaluated. While the literature does not agree one way or the other, this indicates, while present in translation tasks for correct trials [86], they are stronger in incongruent trials [84, 85] and on anomalous endings [74, 75].

4. Discussion

In the identification task false identified pairs elicited a significantly larger N200 and N400 than correct identified pairs for both groups and electrode positions. False identified pairs for group B elicited the largest N200 component. ERN and CRN components are present for comparison processes even before an error occurs [88]. This can be observed in the results of the identification task for Vimmi-German pairs, where a more distinct difference between N200 and N400 can be observed between correct and false identification. The results coincides with previous studies that compared single word processing in German and English, where participants showed a modulation of the N200 [93] or the N200-400 difference between known versus unknown words [94]. It is also indicated, that if memory is easily accessed, the resulting N400 is reduced (ease of effort) [74]. It can be argued that this was the case in this study, since participants had very high scores for translation tasks, especially for translating from Vimmi to German. Previous studies for N400 components showed concrete words caused a larger N400 than abstract words [77, 101, 102].

4.1. Limitations and improvements

Comparison of correct and incorrect answers for the EEG translation paradigms was not possible due to very few wrong answers by participants. This number was further diminished by removed trials due to artifacts. Additional words would have to be learned in the same time span, or the EEG measurement could be done in an earlier stage of learning to increase likelihood of wrong answers.

To increase the number of data points, to get clearer results for ERP curves, the EEG paradigms could have been done repeatedly.

The analysis of the time it took participants to answer in paradigm one (Free entry of German-Vimmi pairs), showed too much variability to effectively investigate the results using ERPs. The moment of translation could not be accurately pinpointed.

Other studies with similar focus opted for a within-subject design [22]. Even with a small number of subjects this could show differences between gesture and no gesture use for abstract word learning. It would compensate for subject variability in learning type or language proficiency. Also 6 learning days instead of only 3 learning units were used, showing improvements over no gesture groups following day 3 [22].

5. Conclusion

In conclusion, the increase in memory performance between the use of gestures and no gestures during learning on 3 days was not significant. The time it takes for the enactment effect to be beneficial is longer than anticipated. For most of the results gesture groups caught up to, or surpassed no gesture groups by day three, indicating that with more learning units, they would achieve higher memory performance, as was the case in other studies [22].

N200 and N400 ERP amplitudes did not differ significantly for translation tasks when only taking correct trials into account. However they did differ significantly when comparing correctly identified with falsely identified German-Vimmi pairs.

These findings can be used to improve the study to obtain better results. Mainly using a within-subject design and increasing the number learning units, to get results over a longer time span.

However it can be said that no negative shift of learning results could be detected for subjects learning abstract words with the aid of gestures over a short period of time. The memory performance for gesture groups indicate increased benefits for additional learning units. This would support the use of gestures in foreign language learning even for abstract words, to keep the learning process congruent.

6. References

[1] Macedonia, M. 2014: Bringing back the body into the mind: gestures enhance word learning in foreign language. Frontiers in Psychology

[2] Yamamoto,Y. (2014). Multidimensional vocabulary acquisition through deliberate vocabulary list learning. System 42, 232–243. doi: 10.1016/j.system.2013.12.005

[3] Piaget,J.(1976).“Piaget’s theory,” in Piaget and His School, eds B. Inhelder, H. Chipman, and C. Zwingmann (Berlin: Springer),11–23.

[4] Tomasello, M.(2005). Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press.

[5] Goldin-Meadow, S., and Alibali, M.W.(2013). Gesture’s role in speaking, learning, and creating language. Annu. Rev. Psychol. 64, 257–283. doi: 10.1146/annurevpsych-113011-143802

[6] Macedonia, M. and von Kriegstein, K. 2012Gestures Enhance Foreign Language Learning, Biolinguistics 6.3–4: 393–416,

[7] Zimmer, H. 2001. Why do Actions speak louder than words. Action memory as a variant of encodingmanipulations or the result of a specific memory system? In H. D. Zimmer, R. Cohen, J. M. J. Guynn, R. Engelkamp & M. A. Foley (eds.), Memory for Action: A Distinct form of episodic memory, 151– 198. New York: Oxford University Press.

[8] Quinn-Allen, Linda 1995. The effects of emblematic gestures on the development and access of mental representations of French expressions. The Modern Language Journal 79, 521–529.

[9] Macedonia, Manuela. 2003. Sensorimotor enhancing of verbal memory through ‘Voice Movement Icons’ during encoding of foreign language (German: Voice Movement Icons. Sensomotorische Encodierungsstrategie zur Steigerung der quantitativen und qualitativen Lerneffizienz bei Fremdsprachen) University of Salzburg: PhD Thesis.

[10] Tellier, M. 2008. The effect of gestures on second language memorisation by young children. Gesture 8, 219–235.

[11] Radonvilliers, C. F. L. D. (1768). De la ManieÌre d’Apprendre les Langues. Paris: Saillant.

[12] Asher, J. J.(1969). The total physical response technique of learning. J. Spec. Educ. 3, 253–262.doi:10.1177/002246696900300304

[13] Engelkamp, J. (1980). Some studies on the internal structure of propositions. Psychol. Res. 41, 355–371.doi:10.1007/BF00308880

[14] Cohen,R.L.(1981).Onthegeneralityofsomememorylaws. Scand. J.Psychol. 22, 267–281. doi:10.1111/j.1467-9450.1981.tb00402.x

[15] Quinn-Allen, L.(1995).The effects of emblematic gestures on the development and access of mental representations of French expressions. Mod. Lang. J. 79, 521–529. doi:10.1111/j.1540-4781.1995.tb05454.x

[16] Kelly, S. D., Barr, D., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. Journal of Memory and Language, 40, 577_592.

[17] Willems, Roel M. & Peter Hagoort. 2007. Neural evidence for the interplay between language, gesture, and action: A review. Brain and Language 101, 278–289.

[18] Willems, Roel M., Aslı Özyürek & Peter Hagoort. 2007. When language meets action: The neural integration of gesture and speech. Cerebral Cortex 17, 2322–2333.

[19] Tellier, M. (2008). The effect of gestures on second language memorization by young children. Gesture 8, 219–235. doi: 10.1075/gest.8.2.06tel

[20] Kelly, S.D., Mcdevitt, T., and Esch, M. (2009). Brief training with cospeech gesture lends a hand to word learning in a foreign language. Lang. Cogn. Process. 24, 313–334. doi: 10.1080/01690960802365567

[21] Macedonia, M., Müller, K., and Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. Hum. Brain Mapp. 32, 982–998. doi: 10.1002/hbm.21084

[22] Macedonia, M., and Knösche, T. R. (2011). Body in mind: how gestures empower foreign language learning. Mind Brain Educ. 5, 196–211. doi: 10.1111/j.1751- 228X.2011.01129.x

[23] Bergmann, K., and Macedonia, M. (2013). “A virtual agent as vocabulary trainer: iconic gestures help to improve learners’ memory performance,” in Intelligent Virtual Agents, Lecture Notes in Artificial Intelligence, eds R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira (Berlin: Springer), 139–148.

[24] Mayer, K.M., Yildiz, I.B., Macedonia, M.,and Von Kriegstein, K. (2014). Motor and visual brain areas support foreign language word learning. Curr. Biol. (in press).

[25] Macedonia, M. (2013). Learning a second language naturally: the voice movement icon approach. J. Educ. Dev. Psychol. 3, 102–116.doi: 10.5539/jedp.v3n2p102

[26] Macedonia, M., Groher ,I., and Roithmayr, F. (2014). Intelligent virtual agents as language trainers facilitate multilingualism. Front. Psychol. 5:295. doi: 10.3389/fpsyg.2014.00295

[27] Engelkamp, J. & H. Krumnacker. 1980. Imaginale und motorische Prozesse beim Behalten verbalen Materials. Zeitschrift für experimentelle und angewandte Psychologie, 511–533.

[28] Engelkamp, J. & H. D. Zimmer. 1985. Motor programs and their relation to semantic memory. German Journal of Psychology, 239–254.

[29] Masumoto, K., M. Yamaguchi, K. Sutani, S. Tsuneto, A. Fujita & M. Tonoike. 2006. Reactivation of physical motor information in the memory of action events. Brain Research 1101, 102–109.

[30] Porter, A. (2012). A helping hand with language learning: teaching French vocabulary with gesture. Lang. Learn. J. 1–21. doi: 10.1080/09571736.2012.750681

[31] Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. J. Verbal Learn. Verbal Behav. 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

[32] Krönke, K.-M., Mueller, K., Friederici, A. D., and Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. Cortex 49, 2553–2568. doi: 10.1016/j.cortex.2012.11.016

[33] Macedonia, M., and Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. Mind Brain Educ. 8, 74–88. doi: 10.1111/mbe.12047

[34] Klimesch, Wolfgang. 1994. The Dtructure of Long-Term Memory: A Connectivity Model of Semantic Processing. Hillsdale, NJ: Erlbaum.

[35] Wig, G. S., M. B. Miller, A. Kingstone & W. M. Kelley. 2004. Separable routes to human memory formation: Dissociating task and material contributions in the prefrontal cortex. Journal of Cognitive Neuroscience 16, 139–148.

[36] Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. Psychon. Bull. Rev. 15, 495–514. doi: 10.3758/PBR.15.3.495

[37] Straube, B., Green, A., Weis, S., and Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: an fMRI study. PLoS ONE 7:e51207. doi:10.1371/journal.pone.0051207

[38] Pulvermuller, F. (1999). Words in the brain's language. Behav. Brain Sci. 22, 253– 279; discussion 280–336. doi: 10.1017/S0140525X9900182X

[39] González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., et al. (2006). Reading cinnamon activates olfactory brain regions. Neuroimage 32, 906–912. doi: 10.1016/j.neuroimage.2006.03.037

[40] Barros-Loscertales, A., Gonzalez, J., Pulvermuller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., et al. (2012). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. Cereb. Cortex 22, 2554–2563. doi: 10.1093/cercor/bhr324

[41] Zimmer, H. D. 2001. Memory for Action: A Distinct Form of Episodic Memory? Oxford: Oxford University Press.

[42] Binder, J. R., C. F. Westbury, K. A. McKiernan, E. T. Possing, & D. A Medler. 2005. Distinct brain systems for processing concrete and abstract concepts. Journal of Cognitive Neuroscience 17, 905–917.

[43] Allen, R. & C. Hulme. 2006. Speech and language processing mechanisms in verbal serial recall. Journal of Memory and Language 55, 64–88.

[44] Fliessbach, K., S. Weis, P. Klaver, C. E. Elger & B. Weber. 2006. The effect of word concreteness on recognition memory. NeuroImage 32, 1413–1421.

[45] Romani, C., S. McAlpine & R. C. Martin. 2008. Concreteness effects in different tasks: Implications for models of short-term memory. Quarterly Journal of Experimental Psychology 61, 292–323.

[46] Pulvermüller, F. 2002. The Neuroscience of Language: On Brain Circuits of Words and Serial Order. Cambridge, MA & New York: Cambridge University Press.

[47] Clarebout, G., Elen, J., Johnson, W. L., and Shaw, E. (2002), “Animated Pedagogical Agents: An Opportunity to be Grasped?, Journal of Educational Multimedia and Hypermedia, 11(3), 267-286.

[48] Moreno, R., Mayer, R., and Lester, J. (2000), “Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of their Interaction,” paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2000. http://www.editlib.org/p/16160

[49] Cassell, J., and Thorisson, K. R. (1999), “The power of a nod and a glance: envelope vs. emotional feedback in animated conversional agents,” Applied Artificial Intelligence, 13(4), 519-538.

[50] Johnson, A. M., Ozogul, G., Moreno, R., and Reisslein, M. (2013), “Pedagogical Agent Signaling of Multiple Visual Engineering Representations: The Case of the Young Female Agent,” Journal of Engineering Education, 102(2), 319-337. doi: 10.1002/jee.20009

[51] Höök, K., Persson, P., and Sjölinder, M. (2000), “Evaluating users' experience of a character-enhanced information space,” AI Communications, 13(3), 195-212.

[52] Cassell, J. (2000), Embodied conversational agents, Cambridge, Mass. ; London: MIT Press.

[53] Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. (1997), “The persona effect: affective impact of animated pedagogical agents,” paper presented at the Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, Atlanta, Georgia, USA.

[54] Buschmeier, H., & Kopp, S. (2011). Towards Conversational Agents That Attend to and Adapt to Communicative User Feedback. In H. Vilhjálmsson, S. Kopp, S. Marsella & K. Thórisson (Eds.), Intelligent Virtual Agents (Vol. 6895, pp. 169-182): Springer Berlin Heidelberg.

[55] Macedonia, M. (2014). A bizarre pedagogical virtual trainer enhances memory for words in a foreign language. International Journal of Computer Science and Artificial Intelligence, 4(2), 24-34.

[56] Macedinia, M., Kern, R., and Roithmayr, F. (2014). Do Children Accept Virtual gents as Foreign Language Trainers? International Journal of Learning, Teaching and Educational Research Vol. 7, No. 1, pp. 131-137, August 2014

[57] Macedonia, M., Bergmann, K., & Roithmayr, F. (2014). Imitation of a pedagogical agent’s gestures enhances memory for words in second language.

[58] Muzzio, I. A., Kentros, C., and Kandel, E. (2009). What is remembered? Role of attention on the encoding and retrieval of hippocampal representations. J. Physiol. 587,2837–2854. doi: 10.1113/jphysiol.2009.172445

[59] Aly,M., and Turk-Browne, N. B. (2016). Attention stabilizes representations in the human hippocampus. Cereb. Cortex 26,783–796. doi: 10.1093/cercor/bhv041

[60] Hauk, O., Johnsrude, I., and Pulvermüller, F.(2004). Somatotopic representation of action words in human motor and premotor cortex. Neuron 41, 301–307. doi: 10.1016/S0896-6273(03)00838-9

[61] Repetto, C., Colombo, B., Cipresso, P., and Riva, G. (2013). The effects of rTMS over the primary motor cortex: the link between action and language. Neuropsychologia 51, 8–13. doi: 10.1016/j.neuropsychologia.2012. 11.001

[62] Repetto, C., Colombo, B., and Riva, G. (2015). Is Motor Simulation Involved During Foreign Language Learning? A Virtual Reality Experiment. Newcastle upon Tyne: SAGE. doi: 10.1177/2158244015609964

[63] Borghi, A. M., Scorolli, C., Caligiore, D., Baldassarre, G., and Tummolini, L. (2013). The embodied mind extended: using words as social tools. Front. Psychol. 4:214. doi: 10.3389/fpsyg.2013.00214

[64] Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., and Tummolini, L. (2016). The challenge of abstract concepts. Psychol. Bull. 143, 263–292. doi: 10.1037/bul0000089

[65] Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., and Del Campo, E. (2011). The representation of abstract words: why emotion matters. J. Exp. Psychol. Gen. 140, 14–34. doi: 10.1037/a0021446

[66] Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., et al. (2014). The neural representation of abstract words: the role of emotion. Cereb. Cortex 24, 1767–1777. doi: 10.1093/cercor/bht025

[67] Berger , H. (1929). Ueber das Elektrenkephalogramm des Menschen. Archives fur Psychiatrie Nervenkrankheiten , 87 , 527 – 570 .

[68] Jasper , H. H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology , 10 , 371 – 375 .

[69] American Encephalographic Society. (1994). Guidelines for standard electrode position nomenclature. Journal of Clinical Neurophysiology , 11 , 111 – 113 .

[70] Walter , W. G., Cooper , R., Aldridge, V. J., McCallum, W. C., & Winter , A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. Nature , 203 , 380 – 384 .

[71] Sutton , S. , Braren , M. , Zubin , J. , & John , E. R. (1965). Evoked potential correlates of stimulus uncertainty. Science , 150 , 1187 – 1188 .

[72] Sanders, LD and Neville, HJ (2003). An ERP study of continuous speech processing. II. Segmentation, semantics, and syntax in non-native speakers. Cognitive Brain Research 15(3): 214–227.

[73] Rossion , B. , & Jacques , C. (2012). The N170: Understanding the time course of face perception in the human brain . In S. J. Luck & E. S. Kappenman (Eds.), The Oxford Handbook of Event-Related Potential Components (pp. 115 – 141). New York : Oxford University Press .

[74] Kutas , M. , & Hillyard , S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science , 207 , 203 – 205 .

[75] Connolly, J, Phillips, N, Stewart, S, and Brake, W (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. Brain and Language 43(1): 1–18.

[76] Van Petten, C and Luka, BJ (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain and Language 97(3): 279–293.

[77] Holcomb, P, Grainger, J, and O'Rourke, T (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. Journal of Cognitive Neuroscience 14(6): 938–950.

[78] Osterhout, L and Holcomb, PJ (1992). Event-related brain potentials elicited by syntactic anomaly. Journal of Memory and Language 31: 785–806.

[79] Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. Trends in Cognitive Sciences, 11, 251_257.

[80] Rugg, M., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. Nature, 392, 595_598.

[81] Schott, B., Richardson-Klavehn, A., Heinze, H. J., & Du¨zel, E. (2002). Perceptual priming versus explicit memory: Dissociable neural correlates at encoding. Journal of Cognitive Neuroscience, 14, 578_592.

[82] Klaver, P., Fell, J., Dietl, T., Schür, S., Schaller, C., Elger, C. E., & Ferna´ndez, G. (2005). Word imageability affects the hippocampus in recognition memory. Hippocampus, 15, 704_712.

[83] Macedonia, M., Müller, K., Friederici, A.: Neural correlates of high performance in foreign language vocabulary learning. Mind, Brain and Education 4(3), 125—134 (2010)

[84] Gehring, W. J., Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Probability effects on stimulus evaluation and response processes. Journal of Experimental Psychology: Human Perception and Performance, 18(1), 198–216.

[85] Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. Psychophysiology, 33(3), 282–294.

[86] Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. Psychological Review, 111(4), 931–959.

[87] Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the "error negativity" specific to errors? Biological Psychology, 51(2–3), 109–128.

[88] Vidal, F., Burle, B., Bonnet, M., Grapperon, J., & Hasbroucq, T. (2003). Error negativity on correct trials: A reexamination of available data. Biological Psychology, 64(3), 265–282.

[89] Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related potential study. Brain Research, 1147, 192_208.

[90] Misra, M., Guo, T., Bobb, S., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. Journal of Memory and Language, 67, 224_237.

[91] Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. Bilingualism: Language and Cognition, 4, 169_178.

[92] Verhoef, K. M. W., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. Cognition, 110, 84_99.

[93] Elston-Gu¨ttler, K. E., Paulmann, S., and Kotz, S. A. (2005). Who’s in control?: Proficiency and L1 influence on L2 processing. Journal of Cognitive Neuroscience, 17: 1593_610.

[94] Conboy, B. T., & Mills, D. L. (2006). Two languages, one developing brain: Event-related potentials to words in bilingual toddlers. Developmental Science, 9, 1_12.

[95] Repetto C, Pedroli E and Macedonia M (2017) Enrichment Effects of Gestures and Pictures on Abstract Words in a Second Language. Front. Psychol. 8:2136. doi: 10.3389/fpsyg.2017.02136

[96] Luck S. J. (2014), An Introduction to the Event-Related Potential Technique, second edition, Massachusetts Institute of Technology, MIT Press ISBN: 9780262525855

[97] Steinhauer K., Connolly J.F., (2008), Event-Related Potentials in the Study of Language, In: Whitaker, H., A., (2010), CONCISE ENCYCLOPEDIA OF BRAIN AND LANGUAGE, Northern Michigan University, Marquette, MI, USA, ELSEVIER, 191-203

[98] Backman, L., Nilsson, L. G., and Nouri, R. K. (1993). Attentional demands and recall of verbal and color information in action events. Scand. J. Psychol. 34, 246–254. doi: 10.1111/j.1467-9450.1993.tb01119.x

[99] Knudsen, E. I. (2007). Fundamental components of attention. Annu. Rev. Neurosci. 30, 57–78. doi: 10.1146/annurev.neuro.30.051606.094256

[100] Pereira,A.,Ellis,J.,andFreeman,J.(2012).Isprospectivememoryenhancedby cue-actionsemanticrelatednessandenactmentatencoding? Conscious.Cogn. 21, 1257–1266.doi:10.1016/j.concog.2012.04.012

[101] West, W. C. & P. J. Holcomb. 2000. Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. Journal of Cognitive Neuroscience 12, 1024–1037.

[102] Levy-Drori, S. & A. Henik. 2006. Concreteness and context availability in lexical decision tasks. American Journal of Psychology 119, 45–65.

[103] King, A. (2016). Pokémon GO for listening and language development. Retrieved from https://avteducationtalk.wordpress.com/2016/07/13/pokemon-go-for-listening-and-language-development/

[104] Singhal, S., Bagga, S., Goyal, P., & Saxena, V. (2012). Augmented chemistry: Interactive education system. International Journal of Computer Applications, 49 (15), 1-5.

[105] Chang, G., Morreale, P. & Medicherla, P. (2010). Applications of augmented reality systems in education. D. Gibson and B. Dodge (Eds), Proceedings of Society for Information Technology and Teacher Education International Conference, 1380-1385. Chesapeake, VA: AACE.

[106] Lee, S.W.Y., Tsai, C.C., Wu, Y.T., Tsai, M.J., Liu, T.C.,& Hwang, F.K. (2011). Internet-based science learning: A review of journal publications. International Journal of Science Education, 33 (14), 1893–1925.]

[107] Yang, Y. F. (2011). Engaging students in an online situated language learning environment, Computer Assisted Language Learning, 24(2), 181-198.

[108] Godwin-Jones, R. (2016). Augmented reality and language learning: From annotated vocabulary to place-based mobile games. Language Learning & Technology 20(3), 9–19. Retrieved from http://llt.msu.edu/issues/october2016/emerging.pdf

[109] Bower M., Howe C., McCredie N., Robinson A., & Grover D. (2014). Augmented reality in education-cases, places, and potentials. Educational Media International, 51(1), 1–15.

[110] Slussareff, M., & Boháčková, P. (2016). Students as game designers vs. 'just' players: Comparison of two different approaches to location-based games implementation into school curricula. Digital Education Review, 29, 284–297.

[111] Dori, Y., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? Journal of the Learning Sciences, 14,2, 243-279.

[112] Klopfer, E., & Squire, K. (2008). Environmental detectives: the development of an augmented reality platform for environmental simulations. Educational Technology Research and Development, 56(2), 203–228.

[113] Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. Journal of Science Education and Technology, 18 (1), 7-22.

[114] Ivanova, M.,& Ivanov, G. (2011). Enhancement of learning and teaching in computer graphics through marker augmented reality technology. International Journal on New Computer Architectures and their applications, 1(1), 176-184.

[115] Yuen, C. S. (2011). Augmented reality (AR) in education. Presentation in Creating Futures through Technology Conference, Mississippi, Retrieved from http://www.slideshare.net/scyuen/augmented-reality-ar-in-education

[116] Rose, F. J., & Bhuvaneswari, R. (2014). Word recognition incorporating augmented reality for linguistic e-conversion. Educational Media International, 51(1), 1–15.

[117] Santos, M. E., Lübke, A., Taketomi, T., Yamamoto, G., Rodrigo, M., Sandor, C., & Kato, H. (2016). Augmented reality as multimedia: The case for situated vocabulary learning. Research and Practice in Technology Enhanced Learning, 11(4), 1–23. Retrieved from http://telrp.springeropen.com/articles/10.1186/s41039-016-0028-2

[118] Solak, E., & Cakir, R. (2015). Exploring the effect of materials designed with augmented reality on language learners' vocabulary learning. Journal of Educators Online, 12(2), 50–72.

[119] Liu, P. E., & Tsai, M. (2013). Using augmented-reality-based mobile learning material in EFL English composition: An exploratory case study. British Journal of Educational Technology, 44(1), 1–4.

[120] Holden, C., & Sykes, J. (2011). Leveraging mobile games for place-based language learning. International Journal of Game-Based Learning, 1(2), 1–18.

[121] Lee, S. H., Choi, J., & Park, J.-il. (2009). Interactive e-learning system using pattern recognition and augmented reality. IEEE Transactions on Consumer Electronics (Vol. 55, pp. 883-890). IEEE. doi:10.1109/TCE.2009.5174470

[122] Mahadzir, N. N. N.,Phung, L. F. (2013). The Use of Augmented Reality Pop-Up Book to Increase Motivation in English Language Learning For National Primary School. IOSR Journal of Research & Method in Education (IOSR-JRME) e-ISSN: 2320-7388, p-ISSN: 2320-737X Volume 1, Issue 1 (Jan - Feb 2013), PP 26-38

[123] Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. Electroencephalogr and Clinical Neurophysiology, 83, 6269.

[124] Boiten F, Sergeant J, Geuze R. Event-related desynchronization: the effects of energetic and computational demands. Electro - encephalogr Clin Neurophysiol. 1992;82:302–309.

[125] Dujardin K, Derambure P, Defebvre L, et al. Evaluation of eventrelated desynchronization (ERD) during a recognition task: effect of attention. Electroencephalogr Clin Neurophysiol. 1993;86: 353–356.

[126] Pfurtscheller G, Lopes da Silva FH. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol. 1999;110:1842–1857.

[127] Elul R. The genesis of the EEG. Int Rev Neurobiol 1972; 15:227-272.

[128] Koshino, Y., & Niedermeyer, E. (1975). Enhancement of rolandic mu-rhythm by pattern vision. Electroencephalogr and Clinical Neurophysiology, 38, 535–538.

[129] Pfurtscheller, G., Pichler-Zalaudek, K., & Neuper, C. (1999). ERD and ERS in voluntary movement of different limbs. In G. Pfurtscheller, & F. H. Lopes da Silva, E_ent-Related Desynchronization. Handbook of Electroencephalogr and Clinical Neurophysiology, Re_ised Edition, vol. 6 (pp. 245–268). Amsterdam: Elsevier.

[130] Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. Annual Re_iew of Physiology, 55, 349–374.

[131] Kalcher J, Pfurtscheller G. Discrimination between phase-locked and nonphase- locked event-related EEG activity. Electroenceph clin Neurophysiol 1995; 94:381-483.

[132] Steriade, M., Gloor, P., Llinas, R. R., Lopes da Silva, F. H., & Mesulam, M. M. (1991). Basic mechanisms of cerebral rhythmic activities. Electroencephalogr and Clinical Neurophysiology, 76, 481–508.

[133] Pfurtscheller G, Lopes da Silva FH. Event-related desynchronization and related oscillatory phenomena of the brain, Handbook of electroencephalography and clinical neurophysiology, vol. 6, revised edition. Amsterdam: Elsevier, 1999.

[134] Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Rev. 1999;29:169–195.

[135] Klimesch W, Pfurtscheller G, Schimke H. Pre- and poststimulus processes in category judgment tasks as measured by event-related desynchronization (ERD). J Psychophysiol. 1992;6:186–203.

[136] Bastiaansen MCM, Böcker KBE, Brunia CHM. Event-related desynchronization during anticipatory attention for an upcoming stimulus: a comparative EEG/EMG study. Clin Neurophysiol. 2001;112:393–403.

[137] Krause MC, Lang AH, Laine M, et al. Event-related EEG desynchronization and synchronization during an auditory memory task. Electroencephalogr Clin Neurophysiol. 1996;98:319–326.

[138] Steriade M, Llinas R. The functional states of the thalamus and the associated neuronal interplay. Phys Rev. 1988;68:649–742.

[139] Klimesch W, Doppelmayr M, Russegger H, et al. Theta band power in the human scalp EEG and the encoding of new information. Neuroreport. 1996;7:1235–1240.

[140] Sterman MB, Kaiser DA, Veigel B. Spectral analysis of event-related EEG responses during short-term memory performance. Brain Topogr. 1996;9(1):21–30.

[141] Röhm D, Klimesch W, Haider H, et al. The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. Neurosci Lett. 2001;310:137–140.

A. Figures

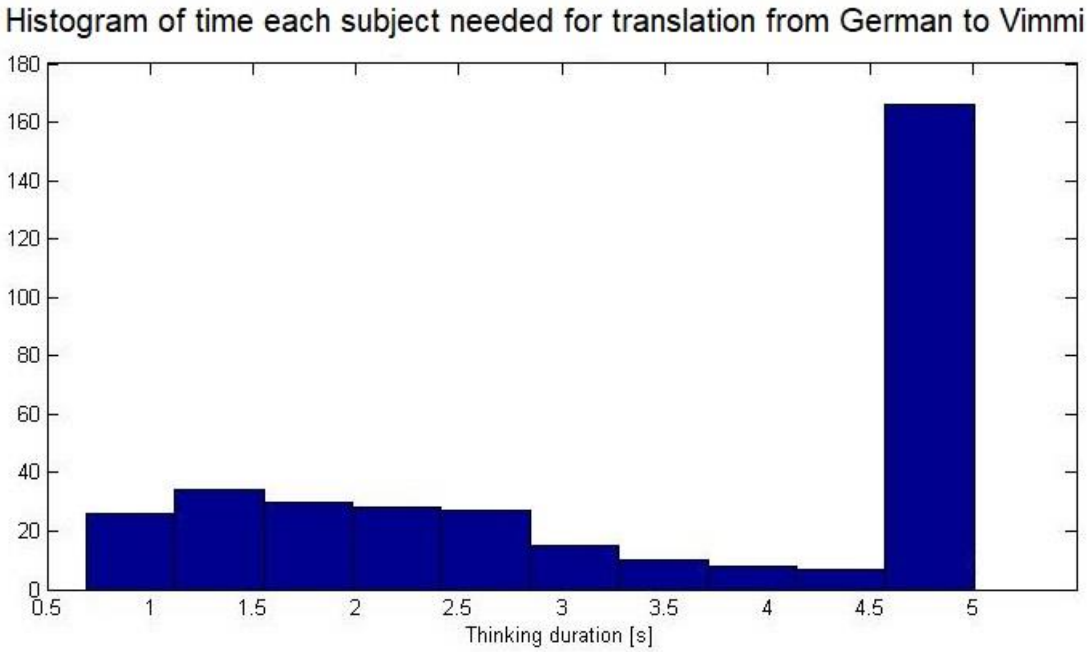


Figure 21: Histogram of response time for translation from German to Vimmi

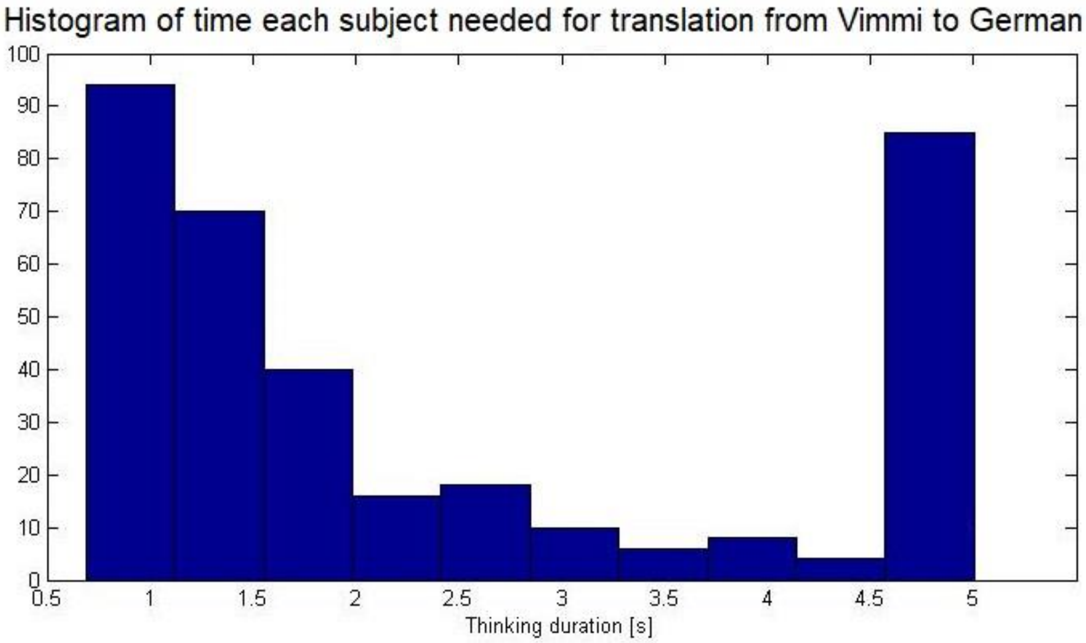


Figure 22: Histogram of response time for translation from Vimmi to German

A. Figures

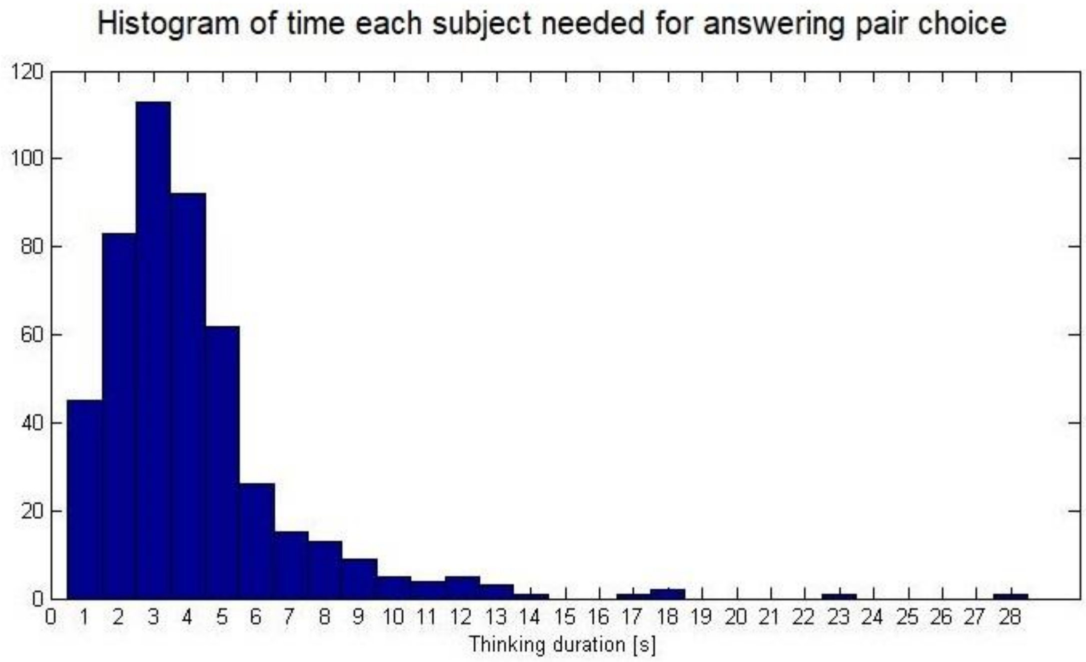


Figure 23: Histogram of response time for answering German-Vimmi pair choices

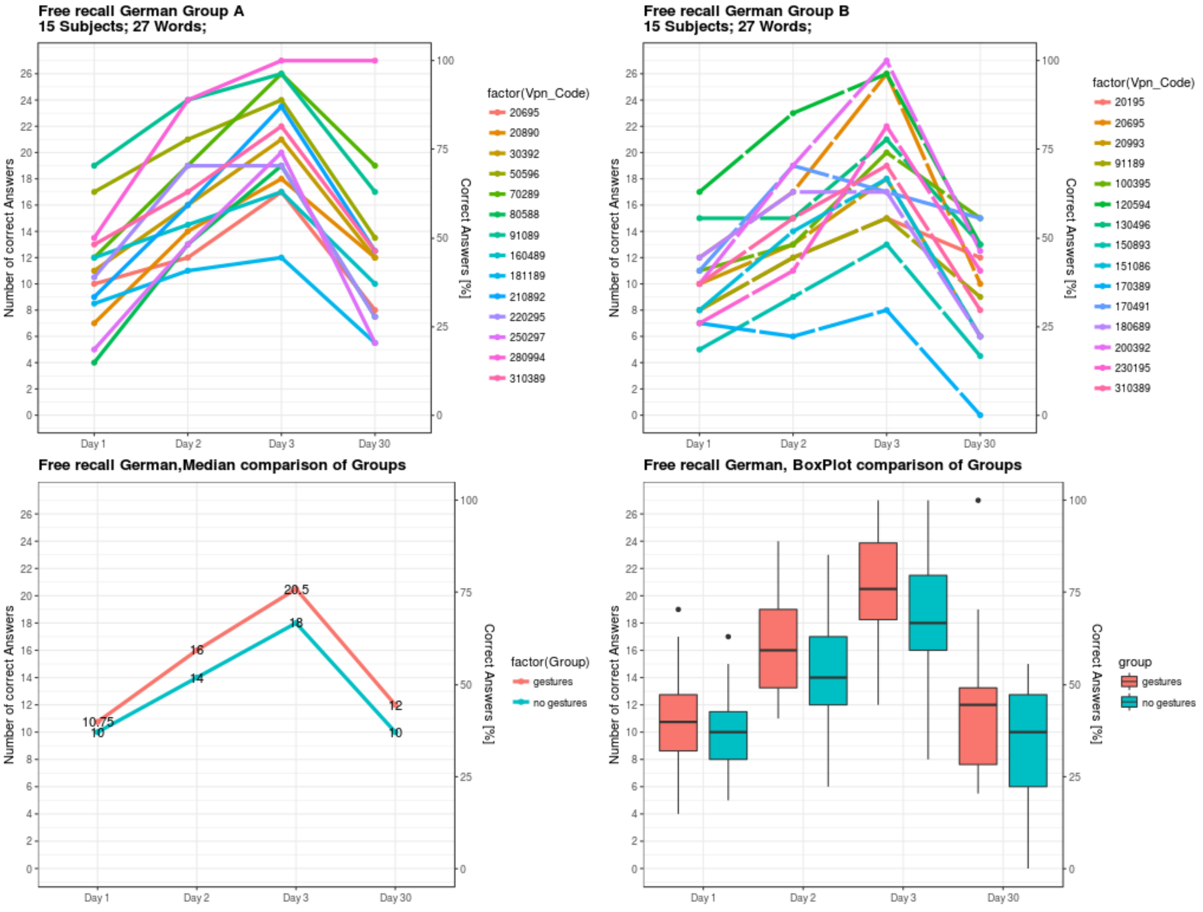


Figure 24: Free recall of German words: Correct answers for group A and B; Median Results for Group A and B; Boxplots for group A and B

A. Figures

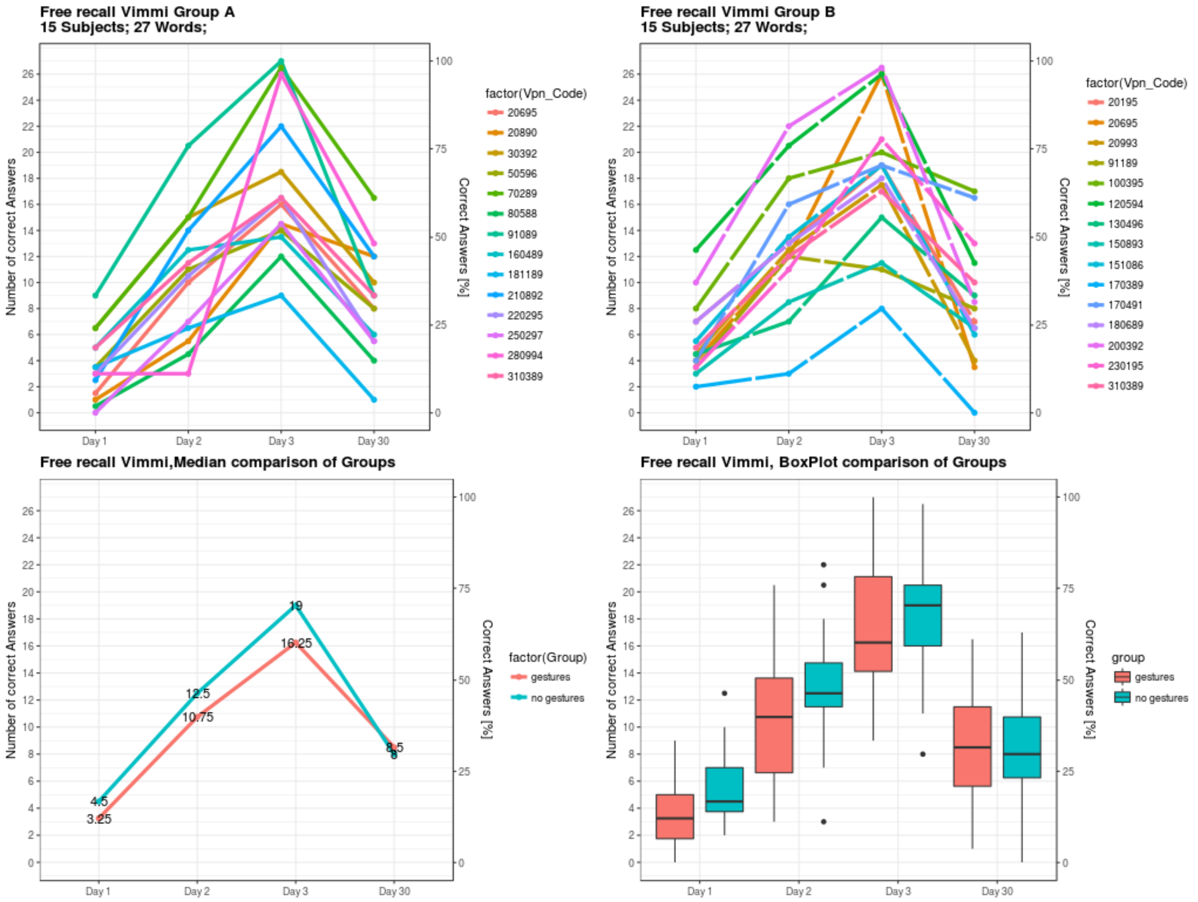
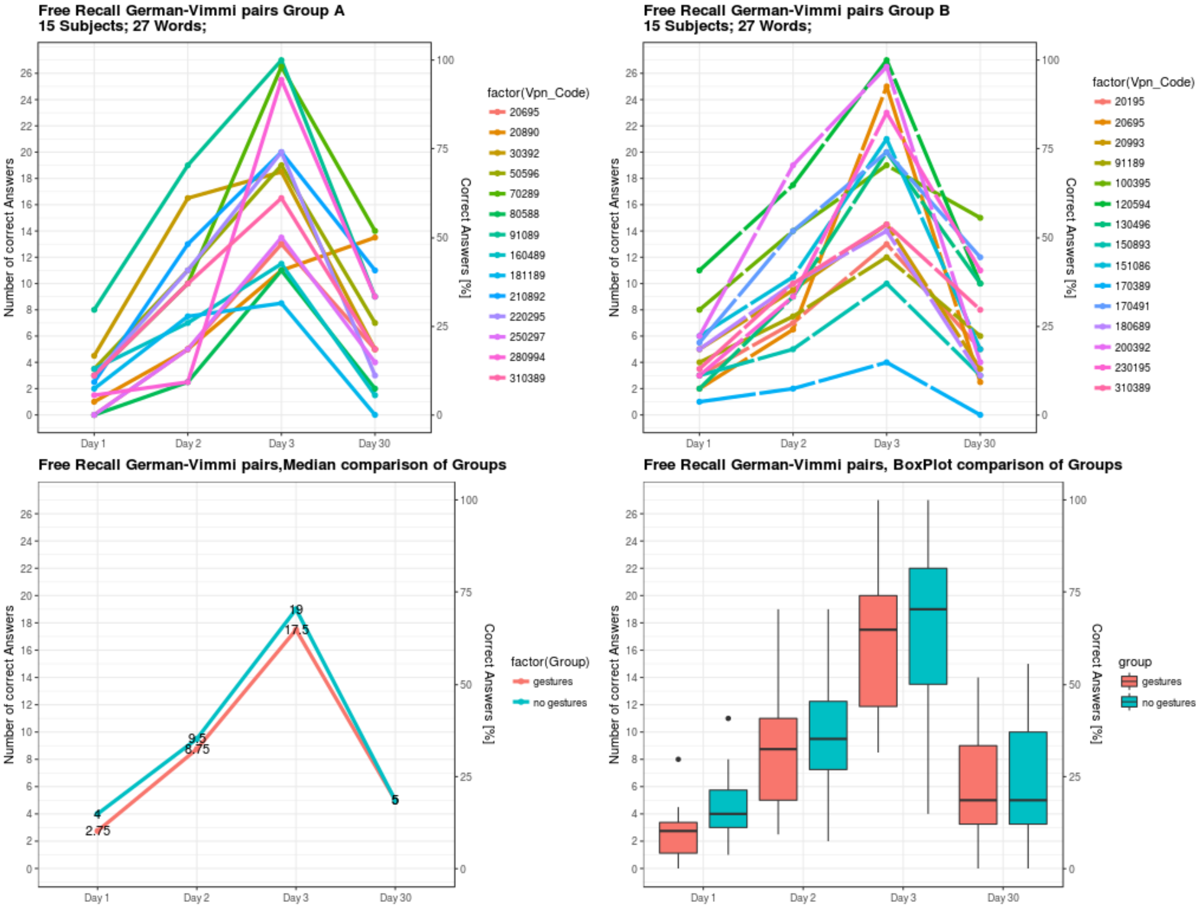


Figure 25: Free recall of Vimmi words: Correct answers for group A and B; Median Results for Group A and B; Boxplots for group A and B

A. Figures



Free Recall German-Vimmi pairs Group A
15 Subjects; 27 Words;

Free Recall German-Vimmi pairs Group B
15 Subjects; 27 Words;

Free Recall German-Vimmi pairs,Median comparison of Groups

Free Recall German-Vimmi pairs, BoxPlot comparison of Groups

Figure 26: Free recall of German-Vimmi pairs: Correct answers for group A and B; Median Results for Group A and B; Boxplots for group A and B

A. Figures

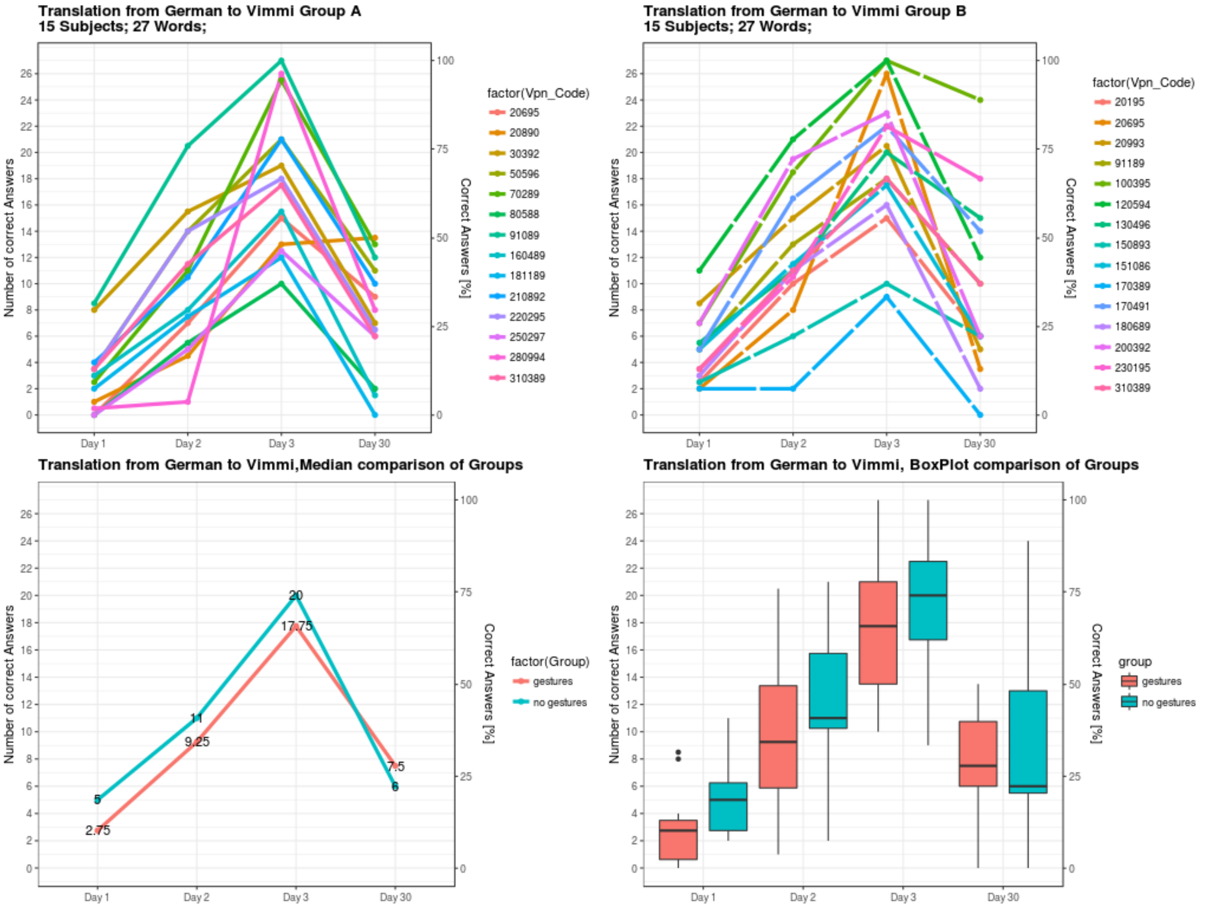


Figure 27: Translation from German to Vimmi: Correct answers results for group A and B; Median Results for Group A and B; Boxplots for group A and B

A. Figures

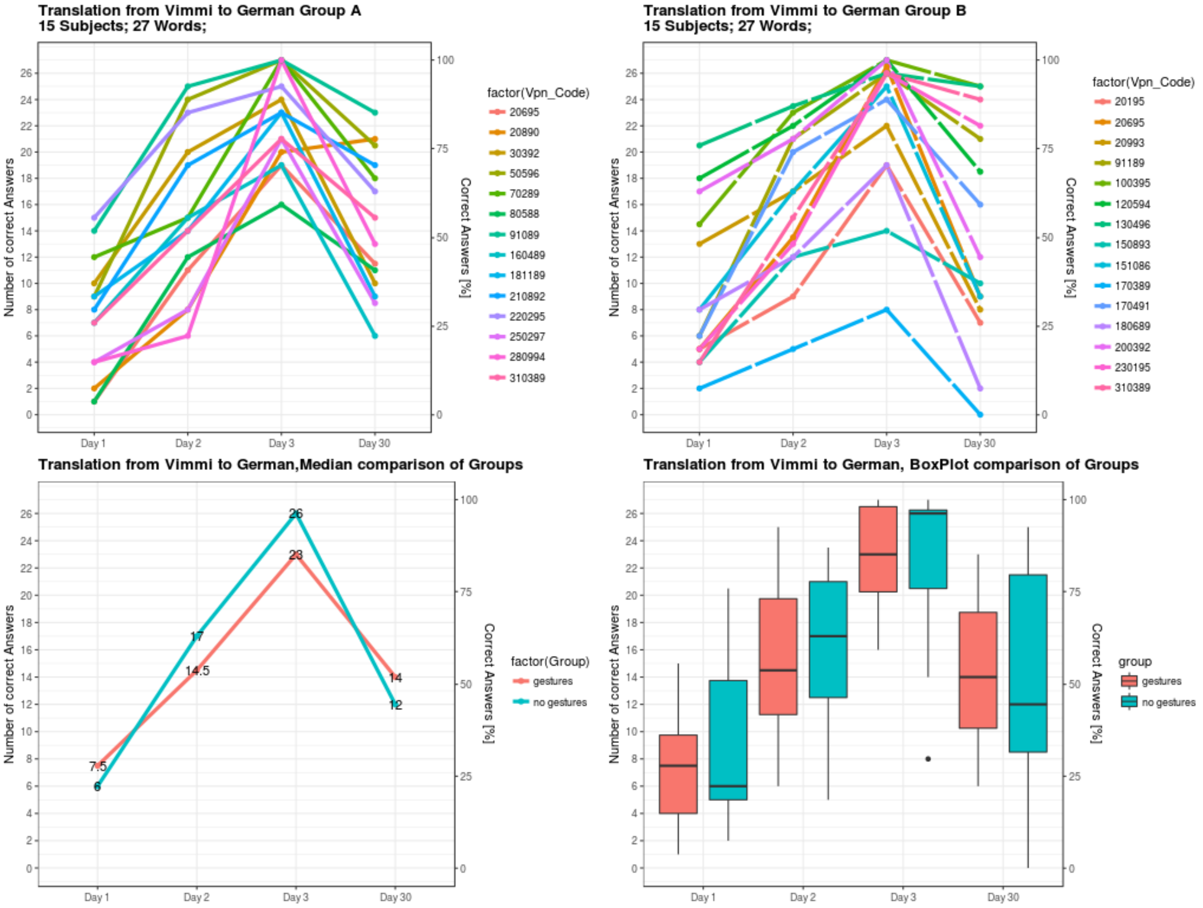


Figure 28: Translation from Vimmi to German: Correct answers results for group A and B; Median Results for Group A and B; Boxplots for group A and B

B. Tables

Table 6: Repeated measures ANOVA of high and low performers Tests of within-subject effect

Source		Sumsquare of type III	df	Mean Square	F	Sig.
TCOND	**Greenhouse-Geisser**	2499,368	2,089	1196,479	60,665	,000
TCOND * GROUP	**Greenhouse-Geisser**	204,452	6,267	32,625	1,654	,148
Error (TCOND)	**Greenhouse-Geisser**	1029,986	52,223	19,723		
TIME	**Greenhouse-Geisser**	20553,724	1,651	12451,725	95,512	,000
TIME * GROUP	**Greenhouse-Geisser**	928,064	4,952	187,411	1,438	,232
Error (TIME)	**Greenhouse-Geisser**	5379,881	41,267	130,368		
TCOND * TIME	**Greenhouse-Geisser**	590,698	5,720	103,270	12,267	,000
TCOND * TIME * GROUP	**Greenhouse-Geisser**	367,590	17,160	21,422	2,545	,001
Error (TCOND*TIME)	**Greenhouse-Geisser**	1203,793	142,998	8,418		

Table 7: Repeated measures ANOVA of high and low performers Tests of between-subject effect

Source	Sumsquare of type III	df	Mean square	F	Sig.
Constant term	62038,621	1	62038,621	524,487	,000
GROUP	2619,885	3	873,295	7,383	,001
Error	2957,108	25	118,284		

B. Tables

Table 8: Repeated measures ANOVA of high performers Tests of within-subject effect

Source		Sumsquare of type III	df	Mean square	F	Sig.
TCOND	**Greenhouse-Geisser**	1281,264	2,253	568,591	26,300	,000
TCOND * GROUP	**Greenhouse-Geisser**	83,321	2,253	36,976	1,710	,197
Error (TCOND)	**Greenhouse-Geisser**	584,614	27,041	21,620		
TIME	**Greenhouse-Geisser**	14191,057	1,599	8872,892	47,762	,000
TIME * GROUP	**Greenhouse-Geisser**	42,186	1,599	26,376	,142	,823
Error (TIME)	**Greenhouse-Geisser**	3565,457	19,192	185,774		
TCOND * TIME	**Greenhouse-Geisser**	581,621	4,994	116,467	10,540	,000
TCOND * TIME * GROUP	**Greenhouse-Geisser**	42,993	4,994	8,609	,779	,569
Error (TCOND * TIME)	**Greenhouse-Geisser**	662,186	59,927	11,050		

Table 9: Repeated measures ANOVA of high performers Tests ofbetween-subject effect

Source	Sumsquare of type III	df	Mean square	F	Sig.
Constant term	43350,914	1	43350,914	333,891	,000
GROUP	80,357	1	80,357	,619	,447
Error	1558,029	12	129,836		

B. Tables

Table 10: Repeated measures ANOVA of low performers Tests of within-subject effect

Source		Sumsquare of type III	df	Mean square	F	Sig.
TCOND	**Greenhouse-Geisser**	1231,442	1,594	772,638	35,945	,000
TCOND * GROUP	**Greenhouse-Geisser**	105,175	1,594	65,990	3,070	,078
Error (TCOND)	**Greenhouse-Geisser**	445,371	20,720	21,495		
TIME	**Greenhouse-Geisser**	7002,143	1,670	4191,968	50,169	,000
TIME * GROUP	**Greenhouse-Geisser**	23,889	1,670	14,302	,171	,806
Error (TIME)	**Greenhouse-Geisser**	1814,424	21,715	83,557		
TCOND * TIME	**Greenhouse-Geisser**	301,126	3,785	79,550	7,228	,000
TCOND * TIME * GROUP	**Greenhouse-Geisser**	21,580	3,785	5,701	,518	,713
Error (TCOND*TIME)	**Greenhouse-Geisser**	541,607	49,210	11,006		

Table 11: Repeated measures ANOVA of low performers Tests of between-subject effect

Source	Sumsquare of type III	df	Mean square	F	Sig.
Constant term	20458,007	1	20458,007	190,092	,000
GROUP	61,807	1	61,807	,574	,462
Error	1399,079	13	107,621		