Check for updates

Depth of Encoding Through Observed Gestures in Foreign Language Word Learning

Manuela Macedonia[1,2]*, Claudia Repetto[3], Anja Ischebeck[4] and Karsten Mueller[5]

[1] Department of Information Engineering, Johannes Kepler University Linz, Linz, Austria, [2] Research Group Neural Mechanisms of Human Communication, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [3] Department of Psychology, Università Cattolica Sacro Cuore, Milan, Italy, [4] Group Cognitive Psychology and Neuroscience, University of Graz, Graz, Austria, [5] Nuclear Magnetic Resonance Unit, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Word learning is basic to foreign language acquisition, however time consuming and not always successful. Empirical studies have shown that traditional (visual) word learning can be enhanced by gestures. The gesture benefit has been attributed to depth of encoding. Gestures can lead to depth of encoding because they trigger semantic processing and sensorimotor enrichment of the novel word. However, the neural underpinning of depth of encoding is still unclear. Here, we combined an fMRI and a behavioral study to investigate word encoding online. In the scanner, participants encoded 30 novel words of an artificial language created for experimental purposes and their translation into the subjects' native language. Participants encoded the words three times: visually, audiovisually, and by additionally observing semantically related gestures performed by an actress. Hemodynamic activity during word encoding revealed the recruitment of cortical areas involved in stimulus processing. In this study, depth of encoding can be spelt out in terms of sensorimotor brain networks that grow larger the more sensory modalities are linked to the novel word. Word retention outside the scanner documented a positive effect of gestures in a free recall test in the short term.

Keywords: word learning, word representation, depth of encoding, foreign language, gesture, memory, fMRI

INTRODUCTION

Vocabulary acquisition in a foreign language (L2) is a tedious and time-consuming task that learners usually perform by reading and repeating the words in bilingual lists. Better results in memorization are achieved if words are enriched by pictures (Mayer et al., 2015) and even better by gestures (for reviews, see Macedonia and Von Kriegstein, 2012; Macedonia, 2014). The positive effect of gestures on memory for words and phrases in native language – the *enactment effect* – has been investigated extensively since the 1980s (for reviews, see Nilsson, 2000; Zimmer, 2001) and has been explained by the following different and partially controversial accounts. First, self-performed action leads to the formation of a *motor trace* that drives the enhancement (Engelkamp and Krumnacker, 1980; Engelkamp and Zimmer, 1985; Nilsson et al., 2000; Masumoto et al., 2006; Eschen et al., 2007; Macedonia et al., 2011); second, representing the word by a meaningful gesture produces a kinematic image of the concept that matches an internal representation of the word (Saltz and Donnenwerthnolan, 1981; Kelly et al., 2009; Macedonia et al., 2011); third, multisensory

processing during gesture performance increases perception and attention and thereby strengthens memory (Kormi-Nouri, 1995, 2000). Studies attributing the enactment effect to the above reasons additionally address depth of encoding as the factor leading to the memory enhancement (Quinn-Allen, 1995; Tellier, 2008; Kelly et al., 2009; Macedonia et al., 2011; Macedonia and Klimesch, 2014).

Depth of encoding was originally described in Craik and Lockhart's *levels-of-processing* (LOP) framework (Craik and Lockhart, 1972; Craik and Tulving, 1975). Thereafter memorization consists of three stages: encoding (processing of incoming information), storage (maintenance and representation of the information), and retrieval (recollection of the information for specific purposes) (Atkinson and Shriffin, 1968).

According to the LOP framework, information is processed hierarchically: sensory perception is considered as "shallow" and decays fast. By contrast, semantic processing is considered as "deep". In deep semantic processing, patterns are recognized, meaning is extracted (Ekuni et al., 2011), and information is kept in memory in a durable way. For example, a word in a foreign language that we only hear is likely to be forgotten because encoded shallowly. In a similar way, vocabulary that we read in lists decays within a short time (Macedonia and Klimesch, 2014). In contrast, encoding becomes deep if vocabulary is learned by selecting it for certain features, i.e., when completing a text or doing other exercises that involve semantic processing (Craik and Tulving, 1975).

The LOP framework describes depth of encoding also in another way. Besides semantic processing, Craik and Tulving (1975) asserted that retention is influenced by the sensorial richness with which verbal material is presented. In other words, adding sensory features to a novel vocabulary item enhances its retention. As documented in the study by Mayer et al. (2015), enriching a novel word in L2 by a picture or by a gesture supports its retention compared to acoustic encoding which is considered as shallow whereby gestures proved to be more efficient than pictures.

Interestingly, also enactment research reconducted depth of encoding through gestures to a multi-component system that drives explicit memory (Engelkamp and Zimmer, 1994). This system consists of a "verbal system", for input and output, and other "non-verbal", i.e., sensorimotor systems. When someone encodes a novel word, different systems are engaged. If the novel word is accompanied by an illustration, the visual system, in addition to the verbal one, will process the information (see also Paivio, 2006). Along this line, if a gesture accompanies the word, the motor system will also participate in encoding. Engelkamp and Zimmer (1994) proposed that the different systems create a sensorimotor representation of the word and that this representation is activated during retrieval. In their account, depth of encoding is explained in terms of sensorimotor complexity of information with a particular focus on the motor component present if a new word is encoded through gestures (Engelkamp and Jahn, 2003). According to Engelkamp and Zimmer (1994), the involvement of the motor system would, however, play the major role in memory enhancement.

Learning Words in a Foreign Language Through Gestures and Depth of Encoding

How are gestures that accompany a word in foreign language related to depth of encoding? How can gestures make verbal information processing deep?

Learners process sensory information on multiple levels if they watch somebody enacting a word or a phrase (Engelkamp and Zimmer, 1994; Klimesch, 1994). Furthermore, if learners perform the gestures themselves, they enhance memory (self-performed task effect, see Engelkamp et al., 1994; Mulligan and Hornstein, 2003). In this case, depth of word encoding is induced first sensorily by enrichment, i.e., by perceiving the gesture and, in a second step, by self-performance.

From a semantic point of view, the match between kinematic image and word semantics leads to deeper encoding than reading or hearing the word. Consider a language teacher illustrating the Japanese word *nomu* (drink) (Kelly et al., 2009) and raising his hand as if holding an invisible cup. Learners observing the teacher will make an involuntary match with an internal template, an image they have stored for the semantics of *drink*. This template can vary in its spatial range to a certain extent, i.e., the gesture can be more or less similar to the internal kinematic image (Kelly et al., 2012).

A match between gesture's perception and gesture's representation seems to be present in the learner's mind. Mismatch paradigms have demonstrated that subjects are susceptible to gestures that do not belong to their inventory and/or do not match their internal kinematic representation of a word (Ozyurek et al., 2007; Willems et al., 2007; Holle et al., 2008; Green et al., 2009). In an fMRI study in foreign language, Macedonia et al. (2011) showed that words that were learned with meaningless, hence not matching, gestures activated a network denoting cognitive control, as in Stroop tasks. This implies that learning a word accompanied by a gesture triggers an internal image, an embodied representation of its semantics.

Hence, we reason that gestures lead to deeper encoding on two paths: first they involve multiple sensory and motor systems in their representation. Second, gestures also induce semantic processing.

Neural Underpinning for Depth of Encoding

The neural substrate of depth of encoding is not fully understood. The literature in the field mainly considers two possibilities: the "semantic processing" view (Otten et al., 2001; Nyberg, 2002) connects depth of encoding with increased activity in prefrontal and temporal brain areas on the base of the task the subject performs (Nyberg, 2002). An incidental task, for example, the detection of the letter A in a word, recruits sensory (visual) regions. The encoding is shallow and the memory performance is less optimal correspondingly. A task demanding more semantic elaboration, like classifying words as for example living or non-living, engages the prefrontal cortex including the core language regions (BA 10, 45, 46, 47). Similarly, deep elaboration tasks additionally involve left temporal regions

(para-hippocampal/fusiform gyri). According to Nyberg (2002), increased activity in these regions is responsible for deeper encoding and leads to better retrievable memory traces.

The other view of depth of encoding attributes the beneficial effects on memory to the recruitment of multiple cortical areas, including motor regions (Klimesch, 1987, 1994; Engelkamp and Zimmer, 1994; Plass et al., 1998; Shams and Seitz, 2008; Shams et al., 2011). In a recent brain imaging study, Macedonia and Mueller (2016) have advanced the hypothesis that depth of encoding through gestures can be connected with the involvement of procedural memory in word learning.

When enriched by gestures, word encoding can potentially exploit both semantic processing and complex sensorimotor perception. These processes need not be mutually exclusive; rather, they could complement each other. Hence, depth of encoding can be an additive process: from shallow sensory perception, to complex sensory perception, and finally to semantic processing.

The Present Study

Studies conducted to date on the effect of gestures on memory have imaged brain responses after participants had learned the words, i.e., after encoding and maintenance of information in memory in the short and long term (for reviews, see Macedonia and Von Kriegstein, 2012; Macedonia, 2014). Studies so far have investigated retrieval but not encoding itself. These studies have not disentangled the benefit of multisensory encoding from the benefit of sensorimotor training when subjects perform themselves the gestures. In fact, it is conceivable that depth of encoding is given by the sensorimotor repetition itself during learning and not necessarily already in the phase of encoding. On the other side, it is also conceivable that depth of encoding is created in the very first steps of perception. Thus, the way learners encode information can have an impact on performance. If we learn vocabulary by reading, or by watching a trainer performing a gesture that is semantically related to the word, we encode the information, respectively, in a shallow or in a deep way.

In L2 education, gestures accompanying words can be used in a two ways:

(a) In encoding, as a *presentation tool* first performed by the teacher while introducing orally novel texts (Macedonia, 2013);

(b) In training, as a *tool after* presentation; in this phase, learners repeat actively the word and the gestures paired to it.

In the present study, we explore encoding online with gestures as a *presentation tool*. We ask if sensorimotor encoding can enhance behavioral performance compared to reading and reading and hearing a word from the very first moment. The question is relevant because in practice, this knowledge could change the way novel vocabulary is introduced to learners. From a neuroscientific point of view, encoding the word from the very first presentation together with sensorimotor information might lead to the creation of extended networks that store information more efficiently and make it better retrievable. These networks

could process the information in a more powerful way than networks consisting of less components (Klimesch, 1994). The outcome of our investigation can be relevant to instruction in order to make vocabulary learning more efficient than it is at present.

The present study wants to shed light on the neural substrate of word encoding with procedures that range from shallow to deep. To our knowledge, our study is the first to measure neural activity during first-stage encoding of words of a foreign language with modalities that grow in complexity and can be connected to different learning procedures in foreign language instruction.

By means of fMRI, we investigate online encoding of novel words through reading (V), audiovisually, i.e., reading and listening to an audio file (AV), and by enriching audiovisual encoding through observed gestures (SMO) with an actress performing a gesture semantically related to the word. These three modalities go from shallow to deep (Klimesch, 1987).

We thus explore the following research questions:

(1) How is stimulus complexity mapped into brain networks?

(2) We hypothesize that complexity of the perceived stimulus is translated into the network extension. As we increment stimulus modalities, functional networks grow larger and from shallow they become deep. In the gesture condition, networks include sensorimotor areas. We thus plan to contrast hemodynamic images of sensorimotor encoding (SM) vs. audiovisual encoding (AV) and audiovisual encoding (AV) vs. reading (V). Also, we will contrast reading (V) with the baseline silence (S).

(3) Is network extension associated with depth of encoding and consequently word retention? We hypothesize that a larger network and particularly a network including sensorimotor areas will impact the depth of encoding. The behavioral outcome of this larger network, compared to less extended networks, will be enhanced word retention.

(4) Do different sensory stimuli involve semantic areas differentially? It is conceivable that with growing depth of encoding, semantic areas are more strongly engaged. In other words, we expect that encoding through observed gestures engages semantic areas more than encoding through reading (baseline) and encoding through reading the word and listening to it.

MATERIALS AND METHODS

Participants

Thirty six right-handed German natives, students of the University of Graz took part in the fMRI experiment. Three subjects had to be excluded for technical reasons, and two failed to complete the experiment. This left 31 participants (thereof 20 females) for analysis with a mean age of 24.35 years, *SD* 3.04. All participants had normal hearing status and normal or corrected vision. They reported no history of neurological or psychiatric illness. They were recruited from the data base of the Psychology Department of the University of Graz. Participants received 10€ compensation for their participation and gave written informed

FIGURE 1 | Screenshots of a video with the actress while performing the gesture for the word "stair".

consent prior to the experiment. The study was approved by the Ethics Committee of the University Graz (Austria).

Stimulus Material

The training materials consisted of 30 three-syllable words in an artificial corpus (**Table 1**) called Vimmi (Macedonia et al., 2011). Vimmi words were randomly generated by a Perl script according to Italian phonotactic rules. In order to avoid associative learning, we controlled the items for similarity to words in European languages previously learned by the subjects, similarity to proper names of products available on the Austrian and German markets, tautological occurrence of syllables, appearance of strings sounding unusual to German-speaking subjects, and high frequency of particular consonants or vowels. We did this to exclude distinctiveness (Hunt and Worthen, 2006) and bizarreness effects (Baddeley and Andrade, 2000) that might have an impact on word retention. We assigned 30 concrete nouns (everyday objects as *knife*, *chair*, *flower*, etc.) and German translations for the Vimmi words arbitrarily. The German nouns were controlled for their frequency of occurrence (familiarity) according to the Vocabulary Portal of the University of Leipzig[1].

The 30 Vimmi words were spoken and recorded by a female German speaker, with each audio file having a length of approximately 1 s. We also recorded 10 videos clips for the sensorimotor (SMO) condition of an actress performing a gesture associated with the German translation of each of 10 Vimmi words (see an example in **Figure 1**), with an average duration of 4.7 s.

fMRI Experiment: Word Encoding in the Scanner

Stimulus presentation and scanner triggering were controlled by a computer outside the scanner room using the software Presentation (version 16.0, Neurobehavioral Systems[2]). Participants lay supine in the scanner. The written Vimmi words,

their translation in German, and the videos were presented via an LCD projector onto a back-projection screen mounted at the participant's feet. Participants viewed the contents of the screen over a mirror mounted on top of the head coil. The Vimmi audio files were presented using an in-ear headphone system (Earplug, NordicNeuroLab AS, Norway).

Procedure

The training materials (30 Vimmi words and their translations as shown in **Table 1**) were presented to the subjects under three conditions:

(1) Visual (V) condition. Written word in Vimmi with German translation (items 1–10);
(2) Audiovisual (AV) condition. Written word in Vimmi with German translation and acoustic presentation of the Vimmi word (items 11–20);
(3) Sensorimotor observation (SMO) condition. Written word in Vimmi with German translation, acoustic presentation of the Vimmi word, and video with an actress performing an iconic gesture semantically related to the word (items 21–30).

Subjects were instructed to memorize the Vimmi words together with their German translations. Subjects were informed that they would perform retention tests outside the scanner after the encoding phase.

The experiment began in the scanner with the instructions shown for 30 s. After a 10-s black screen, the first block started. The blocks for the different conditions were presented in a randomized order. In each trial, a written word in Vimmi appeared at the lower part of the screen with its German translation underneath. It remained visible for 7 s, i.e., V condition. In the AV condition, an audio file played with the onset of the written stimulus. In the SMO condition, a video of the gesture augmented the written and audio stimuli. Each item was presented for 7 s, with an inter-stimulus interval of 10 ± 4 s that varied randomly in 500 ms steps. Due to the repetition time of 2130 ms (for details, see the following chapter), three functional

volumes were completely acquired during stimulus presentation. During the inter-stimulus interval (i.e., in the time between the presentation of the stimuli), a blank screen appeared.

The items were divided into blocks of 10 items, presented separately for each of the three learning conditions. Each block was shown three times, with the 10 items for each condition randomized within the block, for a total of nine blocks and 90 trials. Altogether, every vocabulary item was presented three times in its respective learning condition. The entire encoding part of the study lasted approximately 25 min (**Figure 2**).

MRI Data Acquisition

Imaging was performed using a 3T whole-body system Siemens Trio scanner with an echo-planar capable gradient system and a Siemens-issued 32-element coil. For the functional experiment, we used a T2*-weighted echo-planar imaging (EPI) sequence (flip angle 90°, repetition time 2130 ms, echo time 31 ms,

TABLE 1 | List of stimuli.

Cond. 1 Visual (V)

	Vimmi	German	English
1	Nelosi	Reissverschluss	Zip
2	Gelori	Ohrring	Earring
3	Miruwe	Pfeffermühle	Pepper mill
4	Gepesa	Besen	Broom
5	Mebeti	Becher	Cup
6	Atesi	Treppe	Stairs
7	Lofisu	Foen	Hair dryer
8	Serawo	Giesskanne	Watering can
9	Siroba	Seife	Soap
10	Botufe	Taschentuch	Handkerchief
Cond. 2 Audiovisual (AV)			
11	Suneri	Geige	Violin
12	Wugezi	Regal	Shelf
13	Mewima	Stempel	Stamp
14	Guriwe	Faden	Thread
15	Sigule	Tempel	Temple
16	Lifawo	Stuhl	Chair
17	Bekoni	Kaffee	Coffee
18	Dafipo	Huegel	Hill
19	Pirumo	Erde	Earth
20	Giketa	Blume	Flower
Cond. 3: sensorimotor (SM)			
21	Magosa	Shampoo	Shampoo
22	Uladi	Pullover	Pullover
23	Dirube	Zettel	Sheet of paper
24	Ganuma	Messer	Knife
25	Nabita	Welle	Wave
26	Mesako	Telefon	Telephone
27	Midaro	Spiegel	Mirror
28	Raone	Fernbedienung	Remote control
29	Motila	Banane	Banana
30	Nukile	Poster	Poster

Words used during encoding in the scanner, their translation into German for the participants, and into English for the readers.

image matrix = 64 × 64, 32 axial slices, in-plane resolution: 3 mm × 3 mm, slice thickness: 3 mm, 1 mm gap), which is sensitive to the brain oxygen level dependent (BOLD) contrast. Due to the variable inter-stimulus interval as described above, the total length of the functional experiment varied slightly between the subjects. On average, 746 functional volumes were used resulting in a total experiment length of 26.5 min. The heads of the participants were stabilized with foam padding. Participants were protected from the scanner noise by earplugs embedded in the in-ear headphone system.

For image registration and normalization, a T1-weighted high-resolution image was acquired prior to the functional images. For each subject, one volume was obtained using a three-dimensional magnetization-prepared rapid gradient-echo (MPRAGE) sequence. Acquisition parameters were chosen using a flip angle of 9°, an echo time of 2.07 ms, an inversion time of 900 ms, a repetition time of 1560 ms, and a bandwidth of 230 Hz/Px. A sagittal slice orientation was chosen with 176 slices and an in-plane field-of-view with 256 mm × 256 mm. The nominal image resolution was set to 1 mm × mm 1 × 1 mm with a final image matrix of 176 × 256 × 256.

MRI Data Analysis

Data analysis and pre-processing were performed with SPM8 (Wellcome Department of Cognitive Neurology, London, United Kingdom). Standard processing included realignment, slice-time correction, and normalization to the Montreal Neurological Institute (MNI) space based on the unified segmentation approach (Ashburner and Friston, 2005). After normalization, the resulting voxel size of the functional images was interpolated to an isotropic voxel size of 3 mm × 3 mm × 3 mm. Images were finally smoothed with a Gaussian kernel of 8 mm FWHM. Statistical analysis was performed on the basis of the general linear model as implemented in SPM8. The motion parameters derived from the realignment procedure were entered into the model as parameters of no interest. A high-pass filter (cut-off frequency 1/120 Hz) was used to remove low frequency drifts. No global normalization was used.

A model with three conditions (visual, audiovisual, and sensorimotor) was used with SPM8 and an event-related design. Here, the delta function of the event onsets (corresponding to the onset of each stimulus presentation) for each condition was convolved with the canonical form of the hemodynamic response and its first temporal derivative. Finally, on a single-subject level, contrast images were generated by computing difference images between parameter estimates. Three types of contrast images were obtained (i) subtracting the visual condition (V) from the implicit baseline silence (S), (ii) subtracting the visual condition (V) from the audiovisual condition (AV), and (iii) subtracting the audiovisual (AV) with the sensorimotor condition (SMO). The contrast images calculated for individual subjects were entered into second level random effects analyses (Friston et al., 1999). Finally, in order to correct for multiple comparisons, a family-wise error (FWE) correction was applied on resulting statistical parametric maps with $p < 0.05$ on the cluster level. Here we used a threshold of a minimum cluster size of 30 voxels. The FWE correction
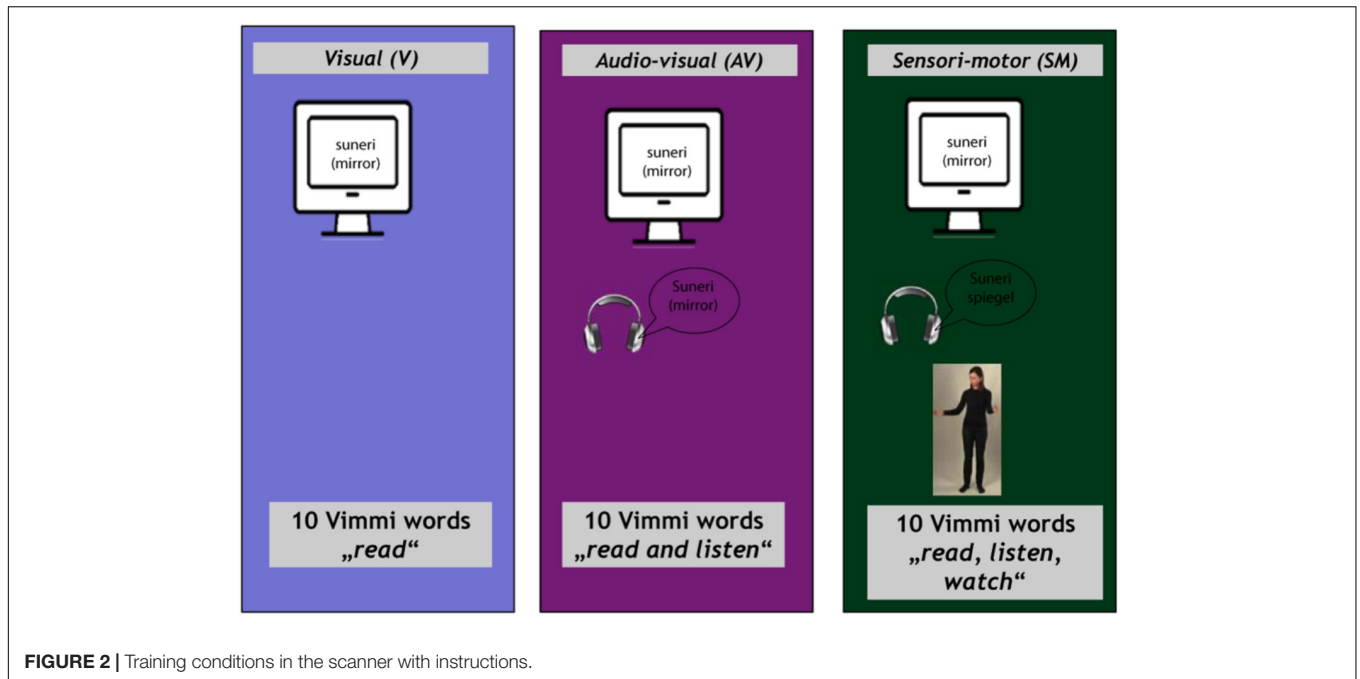
FIGURE 2 | Training conditions in the scanner with instructions.

was performed using SPM8 based on the Gaussian random field theory.

Results

First, we extracted the implicit baseline silence (S) from the visual condition. During the task, participants showed an extended network mapping the reading (written word) and relating it to the word in the native language. It includes the inferior frontal gyrus, the SMA, and the vermis in the left hemisphere. Bilaterally, we found the inferior frontal gyri, the fusiform gyri, the insulae, and the right hippocampus. The right superior temporal pole, the anterior and the posterior cingulate gyri, as well as the superior frontal gyrus participate in the network.

Second, we subtracted the visual (V) from the audiovisual condition (AV). This contrast shows activity in the temporal lobes bilaterally, i.e., right superior temporal gyrus, left middle temporal gyrus, the right inferior frontal gyrus, as well as in the left hippocampus.

Third, we contrasted the audiovisual (AV) with the sensorimotor condition (SMO). Bilaterally, the motor cortices and the inferior parietal lobules were active during encoding. The right superior temporal gyrus, the right vermis, as well as the left fusiform gyrus completed the network.

Areas that were significantly activated during encoding are shown in **Table 2** and **Figure 3**.

Discussion

The results of the fMRI contrast analyses show word processing on the different levels of depth while participants encoded the novel words visually, audiovisually, and by observing the gestures related to the words.

In the first contrast S vs. V, we expected an extended network that maps visual word encoding as well as regions involved in processing the visual written input. We found both, the extended network, and bilaterally the fusiform gyri active in reading tasks and in letter processing (McCandliss et al., 2003). The hippocampus besides memory tasks subserves also functions such as binding and multimodal integration (Breitenstein et al., 2005; Straube et al., 2008). This could be here the case. At a closer look, the activity map reveals the involvement of ventral occipital regions surrounding the fusiform gyrus. This might support the view that reading is accomplished in a ventral occipital temporal network of visual areas (not only focally in the fusiform gyrus), as indicated in a review by Wandell (2011), and more recently by Hannagan et al. (2015). The role of the superior frontal gyrus, here the lower portion of Brodmann area 6, adjacent to primary motor cortices involved in speech articulation, could be that of simulating sound production while participants read the words (Dietz et al., 2005). The left inferior frontal gyrus, fundamental to language tasks (Friederici, 2011; Krönke et al., 2013) in its ventral portion, and to word encoding (Wagner et al., 1998; Kirchhoff et al., 2000) has been claimed to mediate also semantic retrieval (Abutalebi, 2008), and semantic integration between language and gesture (Ozyurek, 2014). This might be the case for our data. In fact, all these areas converge into a universal reading network that orchestrates and reflects the different components of the task (Price, 2012; Rueckl et al., 2015).

During reading, participants learned novel words in Vimmi and connected them with words in their native language. The left middle temporal gyrus, an area engaged in fast mapping of new words (Rodriguez-Fornells et al., 2009), semantic processing (Levy et al., 2004), conceptual semantic representation (Binder and Desai, 2011), and declarative memory (Squire et al., 2004), was also active in the network, together with the anterior and posterior cingulate cortex. The anterior cingulate gyrus, one of the core regions mediating cognitive control, is considered
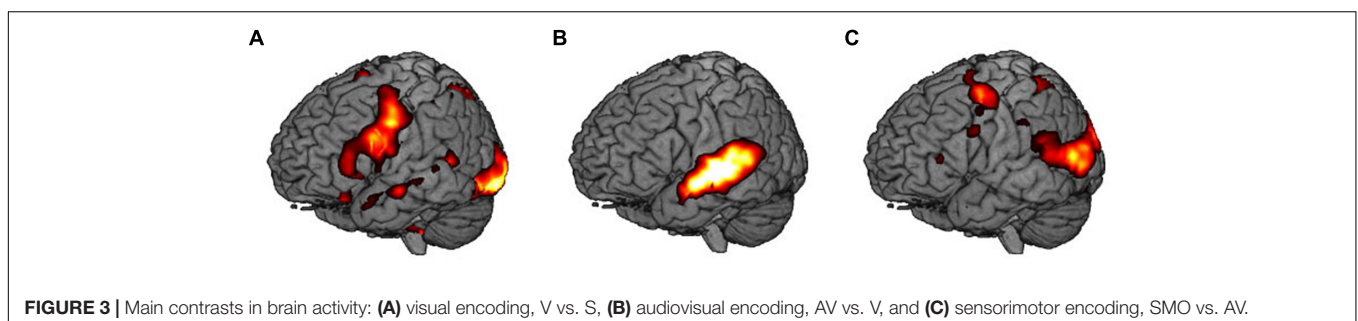
TABLE 2 | Brain areas activated in the different encoding conditions.

Hemisphere		x	y	z	k	Z
	Contrast: baseline – visual (V) vs. silence (S)					
Right	Fusiform gyrus	38	−74	−14	2756	Inf.
Left	Fusiform gyrus	−38	−66	−16	2544	Inf.
Left	Inferior frontal gyrus	−48	14	26	1760	7.55
Left	Cerebellum, vermis	0	−54	−40	141	6.90
Left	SMA	−2	8	56	1253	6.77
Left	Insula	−32	22	−4	830	6.72
Right	Insula	32	28	−2	728	6.60
Right	Hippocampus	22	−30	−2	1549	6.51
Right	Inferior frontal gyrus	46	10	28	932	6.43
Left	Superior parietal lobule	−28	−60	52	739	6.42
Right	Superior parietal lobule	28	−62	52	733	6.25
Right	Superior temporal pole	46	12	−22	123	6.01
Right	Anterior cingulate gyrus	4	−32	34	163	5.70
Left	Middle temporal gyrus	−58	−44	10	89	5.69
Right	Middle temporal gyrus	58	−2	−16	48	5.65
Right	Posterior cingulate gyrus	8	−2	32	46	5.51
Right	Superior frontal gyrus	32	0	66	432	3.93
	Contrast: audiovisual (AV) vs. visual (V)					
Right	Superior temporal gyrus	64	−28	4	3902	Inf.
Left	Middle temporal gyrus	−64	−14	−2	3791	Inf.
Left	Hippocampus	−12	−26	−8	45	5.86
Right	Inferior frontal gyrus	38	16	26	345	4.94
	Contrast: sensorimotor observation (SMO) vs. audiovisual (AV)					
Left	Fusiform gyrus	−22	−82	−10	6389	Inf.
Right	Superior temporal gyrus	58	−40	16	1509	7.07
Left	Precentral gyrus	−36	−4	52	388	6.83
Right	Precentral gyrus	46	2	50	374	6.35
Left	Inferior parietal lobule	−22	−52	54	299	6.25
Right	Inferior parietal lobule	30	−40	52	409	5.60
Right	Cerebellum: vermis	0	−30	−4	56	5.47

Statistical parameter maps were thresholded at p < 0.05 FWE-corrected. Coordinates are reported as given by SPM8 (MNI space). k = cluster size, Z = Z value for the maximally activated voxel of the cluster.

critical to novel word acquisition (Abutalebi and Green, 2007; Raboyeau et al., 2010). As described in a longitudinal fMRI study on L2 word processing (Grant et al., 2015), the anterior cingulate cortex (ACC) interacts with the middle temporal gyrus, a region also present in our data. The ACC detects unfamiliar patterns in the letter sequence and monitors language conflict between native and foreign language (Abutalebi et al., 2012). In a study by

Grant et al. (2015), participants had to judge on first and second language words, as well as interlingual homographs, while lying in the scanner, a task very similar to the one in our study. In the Grant study, the results of region of interest and connectivity analyses showed that regular learning of L2 words changed activation patterns between the ACC and the middle temporal gyrus. While activation decreased in the ACC, it increased in the

FIGURE 3 | Main contrasts in brain activity: **(A)** visual encoding, V vs. S, **(B)** audiovisual encoding, AV vs. V, and **(C)** sensorimotor encoding, SMO vs. AV.

middle temporal gyrus. This indicates that the more familiar the words become, the less the ACC responds. Similarly, the more the words find access to the middle temporal gyrus, the more the latter will react upon word recognition. On this base, our data could also mirror the interaction between ACC and middle temporal gyrus at the very beginning of learning, when subjects were confronted for the first three times with the foreign language words and their translations. The posterior cingulate gyrus (BA 31) might have responded to visual stimuli during reading (Vogt et al., 1992) but also have been involved in processes of episodic memory (Rolls, 2015). In this contrast, we also find activity in the insula, recently detected in word processing (Zaccarella and Friederici, 2015) and engaged in manifold language tasks (for a review, see Oh et al., 2014).

In the contrast AV vs. V, we found the "classical" network for language processing with the participation of the superior and middle temporal gyri essential in auditory word processing (Davis et al., 2009; Segawa et al., 2015), and the right inferior frontal gyrus. In a meta-analysis of neuroimaging studies on the functions of the inferior frontal gyrus, Liakakis et al. (2011) relate this area to fine movement control and working memory. In our study, the right inferior frontal gyrus could be connected to unintentional subvocalization and to memorizing the novel words, as indicated in a similar study by Krönke et al. (2013). Interestingly, in the audiovisual condition, the left hippocampus was significantly more involved than on reading only. This structure has been identified as particularly relevant in word learning, especially during the initial encoding of words (McClelland et al., 1995; Davis and Gaskell, 2009).

In the SMO vs. AV contrast, considering that subjects watched videos with an actress performing gestures, we expected our data to be in line with other studies targeting premotor cortices during movement observation (Buccino et al., 2001; Buccino et al., 2004a,b; Straube et al., 2008; Ogawa and Inui, 2011) and connecting them to the mirror neuron system implicated in action recognition (Rizzolatti, 2015). Despite this, we found bilateral activity in the primary motor cortices and in the parietal lobules. Recently, fMRI studies reported these brain areas also during action observation (Aridan and Mukamel, 2016; Gatti et al., 2016) as belonging to a mirror neuron network active in motor learning tasks that include observation as well as imitation and execution (Sale and Franceschini, 2012). Activations located in the parietal lobules are possibly due to the actress' body observation (Hodzic et al., 2009). In fact, participants processed the actress' image while she performed the gestures. Along this line, activity located in the inferior parietal lobules is possibly due to the actress' body observation (Hodzic et al., 2009). Participant processed in fact her image while she performed the gestures. Furthermore, we found additionally activity in the cerebellar vermis: this brain region might subserve here visual motion processing (Cattaneo et al., 2014).

In the fMRI contrast analyses, we examined the neural activity involved in encoding novel words in a foreign language according to three conditions with growing complexity. The analyses show that the network processing the words also grows in complexity. The network mirrors the modalities added progressively: a basic network engaged in encoding through reading in the visual condition is enlarged by auditory cortices in the audiovisual condition, by motor cortices, and the parietal lobules in the condition with gesture observation. Taken together, the network that encodes novel verbal information by observing gestures seems to connect a set of regions in the left hemisphere: canonical language areas in inferior frontal regions, (pre)motor cortices, and the hippocampi.

Our results are in line with neuroimaging studies that have investigated single word learning (Paulesu et al., 2009; Ye et al., 2011). Despite the different paradigms, these studies have found the recruitment of brain tissue in the left hemisphere, more specifically in the inferior frontal gyrus, the parahippocampal region, and the fusiform gyrus. A study by Straube et al. (2008) explored the retention of Russian sentences accompanied by unrelated or metaphoric gestures by means of a recognition task. Encoding of verbal information with metaphoric gestures, a task similar to the one that our subjects had to accomplish, was associated with neural activity in the left inferior frontal gyrus, the premotor cortex, and the middle temporal gyrus. The hippocampus contributed to retention and correlated with performance. The fMRI study by Straube et al. (2008) yields comparable results for the encoding condition with metaphoric gestures in brain activity with our study for the condition SMO.

Behavioral Experiment: Word Retrieval Outside the Scanner
Procedure
After the fMRI session, participants spent a 5-min break in a room adjacent to the scanner. In this room, subjects completed four different pencil-to-paper tests, as described in the next sections.

Free Recall Tests for German and Vimmi
We gave participants an empty sheet for each language (German and Vimmi) and instructed them to write as many items as possible in each of the two languages (only German or only Vimmi). Participants had 5 min for each of the free recall tests, for a total of 10 min.
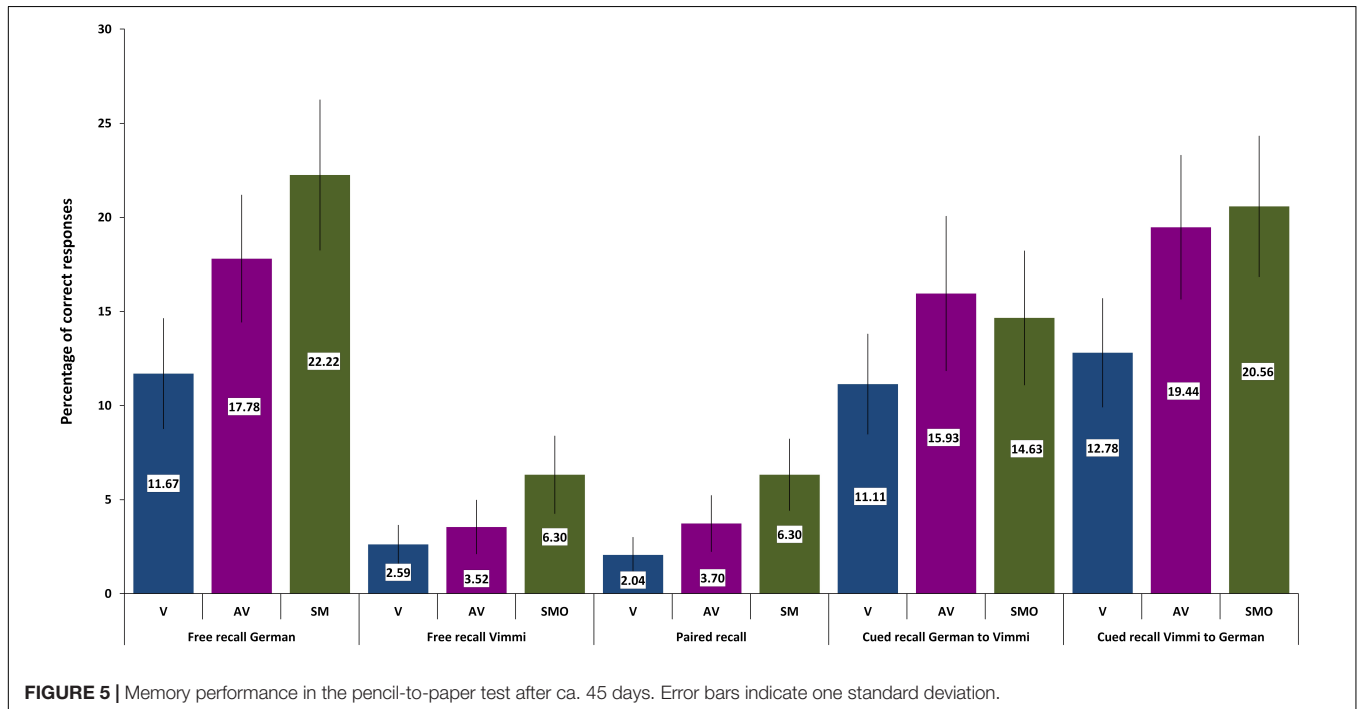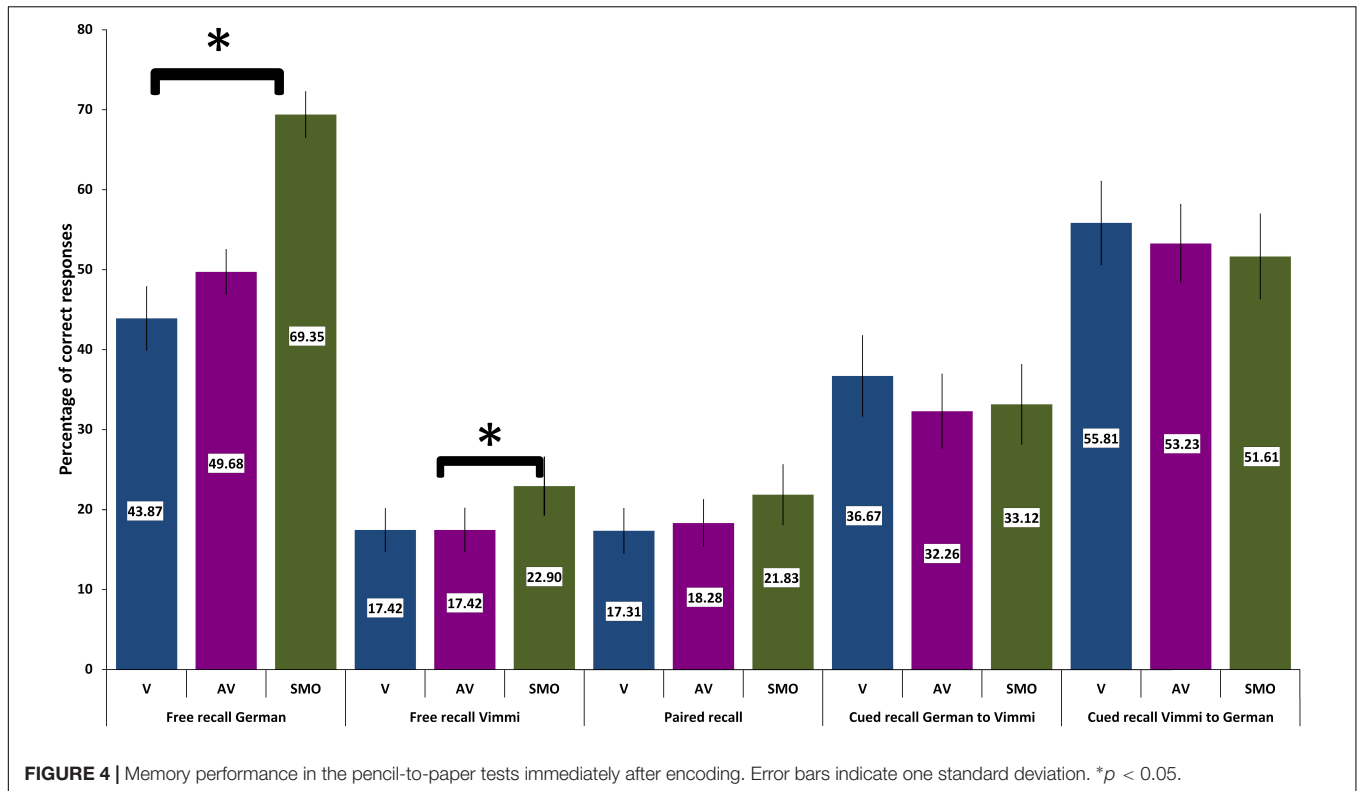
Paired Free Recall Test
Participants were given an empty sheet and instructed to write as many items as possible in both languages (pairs). In this test, items had to be matched (i.e., German and the corresponding word in Vimmi or vice versa). The paired free recall test lasted 5 min.

Cued Recall Tests
We gave participants a randomized list of the 30 trained items to be translated from German into Vimmi (duration 5 min) and then another randomized list of the same words to be translated from Vimmi into German (duration 5 min). The order of the translation from one into the other language alternated from participant to participant.

Follow-Up Tests
We also sought to know how encoding influences decay in time. Therefore, after approximately 45 days, we asked

FIGURE 4 | Memory performance in the pencil-to-paper tests immediately after encoding. Error bars indicate one standard deviation. *$p < 0.05$.

FIGURE 5 | Memory performance in the pencil-to-paper test after ca. 45 days. Error bars indicate one standard deviation.

subjects to participate in a follow-up test. We emailed our participants a link to anonymous word retrieval tests that would document their memory performance. Out of 31 participants, 18 accepted the invitation. The online tests were realized with the Qualtrics software, version 56531, in the Qualtrics Research Suite[3]. Participants completed online the same tests that were administered after encoding, i.e., free recall (German, Vimmi, and paired) and cued recall tests in both translation directions.

---

[3]http://www.qualtrics.com

Each test was time-locked: after a time limit of 5 min, the current test disappeared and the next test was presented. After completion, results contained in the log files were entered into SPSS for behavioral statistical analysis.

Statistical Analyses

For each memory task and for each experimental condition, a performance index was calculated as the percentage of correctly recalled items over the total number of items. Then the performance scores were entered in a repeated measures ANOVA, with the conditions visual (V), audiovisual (AV), and sensorimotor (SMO), as within-subjects factors. *Post hoc* contrasts (within subjects simple contrasts) between conditions were computed when needed. Descriptive statistics results are reported in **Figures 4**, **5** (immediate recall and follow-up performances, respectively).

Finally, in order to assess the influence of the different conditions on neural mechanisms related to depth of encoding, we conducted correlation analyses between brain and behavioral data. We ran two parametric analyses based on the contrast SMO vs. AV. As a regressor for the second level analysis, we entered the behavioral test results showing significant differences among encoding conditions, i.e., (1) the results from the free recall test in German and (2) the results from the free recall test in Vimmi (**Table 3**). The threshold for these two analyses was set to $p < 0.005$ uncorrected, reporting only clusters with a minimal cluster size of 30.

Results

Memory performance

Figure 4 shows that participants recalled more words in German than in Vimmi. For the *free recall* test in *German*, data underlined that performance differed significantly among conditions [$F_{(2,60)} = 28,26$, $p < 0.001$, $\eta^2 = 0.48$]. Sensorimotor encoding proved to be significantly superior to both visual [$F_{(1,30)} = 42.03$, $p < 0.001$; $\eta^2 = 0.59$] and audiovisual encoding [$F_{(1,30)} = 37.92$, $p < 0.001$; $\eta^2 = 0.56$]. In the free recall test in Vimmi, memory performance again differed depending on the type of encoding [$F_{(2,60)} = 3.38$, $p = 0.04$, $\eta^2 = 0.1$], with sensorimotor encoding proving superior to visual [$F_{(1,30)} = 5.23$, $p = 0.03$; $\eta^2 = 0.15$] but not to audiovisual encoding [$F_{(1,30)} = 3.93$, $p = 0.06$].

In the *paired free recall test*, performance did not differ among conditions [$F_{(2,60)} = 1.94$, $p = 0.15$]. In the *cued recall tests from German into Vimmi* and *Vimmi to German*, memory results did not differ with regard to the condition under which subjects had learned [German to Vimmi: $F_{(2,60)} = 0.78$, $p = 0.46$; Vimmi to German: $F_{(2,60)} = 0.76$, $p = 0.47$]. In the follow-up tests, the mean values by condition (**Figure 5**) indicate that overall the recall rates were very low (ranging from about 2 to 22%). Nevertheless, having a look at the descriptive data, interesting differences emerge if we compare the encoding conditions. In *German and Vimmi free recall* and in *paired free recall*, participants scored worst for words encoded by mere reading (V). For words encoded audiovisually (AV), they did better, and best memory results were obtained through sensorimotor encoding. In particular, the difference between the V and SMO conditions in the *German and*

Vimmi free recall tests is worth noting. Participants could retrieve nearly twice as many words encoded with gestures than words learned by mere reading. In the *paired recall test*, participants retrieved three times as many words learned in the SMO than words learned in the V condition. However, due to the high variance in performance among subjects, only the *German free recall* test nearly reached significance [$F_{(2,34)} = 3.19$, $p = 0.054$, $\eta^2 = 0.19$]. In both the cued recall tests, memory performance did not differ among encoding conditions, i.e., from *German into Vimmi* [$F_{(2,34)} = 2.09$, $p = 0.14$] and *from Vimmi into German* [$F_{(2,34)} = 2.64$, $p = 0.09$].

Correlations Between Brain Activity and Behavior

The parametric analysis relating the contrast SMO vs. AV with the mean performance of each single subject in the German free recall and Vimmi free recall tests gave further insights in the neural mechanism of depth of encoding. We hypothesized that the results would parametrically mirror the brain patterns of the contrast, hence possibly involve the motor cortices and/or the parietal lobules or parietal regions involved in semantic encoding.

TABLE 3 | Behavioral average performances in free Vimmi and free German.

Subject	Free German	Free Vimmi
enc_01	63,33	28,89
enc_02	56,67	4,44
enc_03	43,33	18,89
enc_04	80,00	36,67
enc_05	60,00	44,44
enc_07	66,67	25,56
enc_08	53,33	2,22
enc_09	80,00	42,22
enc_10	63,33	3,33
enc_11	50,00	18,89
enc_12	70,00	23,33
enc_15	36,67	13,33
enc_16	46,67	0,00
enc_18	40,00	6,67
enc_19	60,00	20,00
enc_20	23,33	5,56
enc_21	46,67	12,22
enc_22	50,00	16,67
enc_24	63,33	18,89
enc_25	63,33	20,00
enc_26	60,00	11,11
enc_27	50,00	18,89
enc_28	70,00	54,44
enc_29	40,00	5,56
enc_30	60,00	21,11
enc_31	40,00	8,89
enc_32	20,00	0,00
enc_33	56,67	31,11
enc_34	56,67	7,78
enc_35	63,33	54,44
enc_36	50,00	21,11

The correlation with the performance score from the *free recall test in German*, the subjects' native language, shows activity in a number of regions with the biggest area located in the left inferior temporal gyrus. This area is considered to be involved in a large cortical network performing semantic storage (Whitney et al., 2010). During word encoding in an event-related and in a blocked fMRI paradigm, Wagner et al. (1998) measured brain activity. They found that the capacity to memorize verbal information is related to the magnitude of activation in canonical language areas (left prefrontal cortices) as well as in left temporal regions including the inferior temporal gyrus. Similar results were achieved in a study by Kirchhoff et al. (2000) during a semantic task.

The correlation with the performance data from *the free recall test in Vimmi*, the foreign language to the subjects, yielded different results. We found activity in the extra-striate cortex. Also larger areas of activity were present in the left hippocampus and in the left parahippocampal gyrus. These regions are critically involved in memorization tasks and are part of the network for word encoding found in the study by Wagner et al. (1998). We can only speculate on the reason why activity in these areas correlates more with Vimmi word encoding than with German word encoding. Considering that new words in the foreign language need to be stored as sequences of phonemes and graphemes, the hippocampus and the parahippocampal gyrus might perform with larger amplitude because they are more engaged in the Vimmi than in the German task.

Altogether, the correlations between brain activity and behavioral results only partially meet our hypotheses. Expecting a recruitment of semantic areas, we found parametric activity in the left inferior temporal gyrus only for the German words. The Vimmi words recruited canonical memory structures, possibly because of the stronger memory effort needed in order to store grapheme and phoneme sequences in the foreign language.

Discussion

In this combined fMRI and behavioral study, participants encoded 30 words of Vimmi in the scanner according to three conditions (V, AV, and SMO) with 10 words for each condition. The encoding was brief, with three repetitions per word.

We expected that enriching a word with multiple and sensorimotor information behaviorally would enhance its memorability in the short term and in the long term. We obtained significant results in the free recall test in German and Vimmi immediately after encoding. After approximately 45 days, only the free recall test in German nearly reached significance.

In sum, encoding with sensorimotor input led to superior results in the free recall tests in the short term.

It is conceivable that under "normal" conditions, i.e., if learning had occurred outside the scanner, behavioral performance might have matched other studies showing that enriched encoding supports memory for words in L2 significantly. In fact, it is likely that learning in the scanner influences performance. Lying supine inside a noisy "tube" is a novel, uncomfortable, and distracting situation for participants. According to Peelle (2014), noise from the scanner reduces sensitivity to acoustic experimental stimuli, requiring additional resources in executive functions while processing. More significantly, scanner noise can alter both auditory and non-auditory brain networks hence compromise the expected results. In our study, we made use of in-ear headphones. According to the producer, they provide "sufficient" noise insulation and ensure good quality for the presentation of auditory stimuli. The influence of the scanner noise is equally present in all learning conditions (V, AV, and SMO). All conditions are thus potentially equally affected by the scanner noise. We chose a standard EPI sequence with a moderate noise level and not continuous EPI sequences with reduced acoustic noise. An alternative to our analyses could have been a finite impulse response model as described in Peelle (2014).

Although subjects were exposed to only three repetitions, in order to separate encoding from storage, from a LOP perspective, retention was high if compared to similar studies. In fact, Macedonia and Knösche (2011) trained participants on a similar experimental design for 4 days, 3 h per day. Results became significant only on days 3 and 4, after a very high number of repetitions. Interestingly, also in that study, subjects performed best in German free recall at all time points. This might be the case because German free recall is the easiest task to accomplish. Once a concept's semantics is clear, learners already have the "label" for the concept (the word). Learners need not memorize novel sound sequences that do not match syllable structure and sounds of the native language. In that study, after 45 days, in all free recall tests, words encoded with observed gestures scored better than in the other conditions and memory results did not mirror those produced immediately after encoding. In other words, even if in the present study the exposure to stimuli was passive and for a very low number of repetitions, results confirm that sensorimotor encoding is crucial for memory processes.

In our study, contrarily to the majority of the studies in the field, subjects only observed and did not enact the items to be learned. Since the beginning of enactment research, it has been known that self-performing the gestures leads to better memory results than only observing them. This was documented in a seminal study by Cohen (1981), in which the author called the effect unambiguously *subject-performed task effect*. At that time, this issue applied to memory for words in L1. In L2 learning, Macedonia and Klimesch (2014) compared the retrieval of 18 words encoded audiovisually and retrieval of 18 words learned with gestures over 14 months. Although the number of repetitions was also very low (eight repetitions on day 1 and four repetitions on day 8, altogether 12 repetitions) memory performance was highly significant for words encoded through gestures at all times. Particularly impressive were the results after 14 months. Whereas participants had nearly forgotten all words encoded audiovisually (mean retrieval performance 1.15%) participants still could recall 10.45% of the words encoded with self-performed gestures. Although speculative, we reason that best memory results can be achieved if learners encode by watching the gestures and thereafter perform the gestures by themselves. To shed more light on this aspect, further studies should be conducted in order to confirm the possible connection between encoding and active processing.

Correlations Between Brain Activity and Behavior

We considered the SMO condition to allow more depth of encoding. The contrast analysis SMO vs. AV showed that encoding with gestures elicited greater signal intensity in the motor cortices bilaterally and in the parietal lobules. In this line, for better memory performance, we expected more engagement of the motor system. Instead, the correlation analyses indicated greater activity in language and memory areas. Similarly, in a study by Macedonia et al. (2010), higher memory performance by means of gestures was hypothesized to be related to engagement of the motor system. However, correlation analyses between significant brain activity and behavioral performance revealed in that study activity in the left angular gyrus (BA 39), and in the right extra-striate-cortex (BA 19). The left angular gyrus is described in the literature as a brain area subserving integrative functions in word processing as well as semantic integration (Binder et al., 2009). In the present study, the portion of the extra-striate-cortex revealed by the correlation analyses is considered in its functionality as a posterior extension of the angular gyrus. So, both studies seem to have a connection which should be elucidated with further experiments. At present, we only see the possibility of a stronger functional link between sensorimotor encoding and memory areas in the brain which seems to be stronger than for visual or audiovisual encoding.

CONCLUSION

For the first question posed in the introduction, i.e., on how stimulus complexity is mapped into neural networks, our data show that networks grow incrementally in complexity under the three encoding conditions. The growth reflects the sensorial stimuli processed during encoding. For the SMO condition, the one that we considered to be the most efficient, the network includes bilaterally the primary motor cortices, the cerebellum, and the inferior parietal lobules. The participation of the motor system could indicate that procedural memory is engaged in encoding even if participants only observed the stimuli and did not perform the gesture. This could explain why gestures – even if observed during encoding lead to better memory performance.

The second question, regarding network extension connected to depth of encoding and impact on retention as suggested in Craik and Tulving (1975) and Engelkamp and Zimmer (1994), is partly supported by our behavioral data. The SMO condition is associated with significantly better retention only in the free recall tests in both German and Vimmi in the short term.

Our imaging data answer the third question about the participation of semantic cortices in encoding: the network described in the first contrast silence vs. visual encoding includes the middle temporal gyrus, an area involved in deep semantic processing. However, the literature does not describe the V condition leading to particular semantic elaboration of the word.

Furthermore, our data show that reading a word does not have a positive impact on retention as AV and SMO encoding. Here, engagement of the temporal areas might reflect the bilingual presentation in the scanner, and possibly the semantic processing of words in the native language and/or connecting words in L2 to words in L1 (Grant et al., 2015). This is to say that the V condition activated semantic cortices and not the SMO condition as expected. It is contradictive to our expectations and it does not provide evidence that gestures if only observed can engage brain tissues engaged in semantic processing. On the other side, it must be considered that it could be very likely that we do not see the activity in semantic cortices for SMO because the SMO was isolated by subtracting the AV from it. In other words, activity in semantic cortices could have been removed by the fMRI analysis. This is one limitation in the subtraction method: information on shared processing goes lost.

The correlation analysis between the brain data related to the SMO condition and the behavioral free recall tests showed the involvement of the left inferior temporal gyrus, a semantic area active in word encoding, for German words. Vimmi words engaged parametrically more the hippocampus and the parahippocampal gyrus.

In sum, our study suggests that observation of a gesture connected to a word in L2 can make its encoding deeper than only reading the word or reading and listening to it. This seems to be due to the engagement of complex sensorimotor networks more than to deep processing in semantic areas. Complex sensorimotor encoding is associated with better memory results than procedures involving fewer senses and not engaging the motor system.

Hence, although traditional bilingual education mainly employs listening and comprehension activities during encoding of novel verbal material, it could be fruitful to introduce sensorimotor enrichment in L2 word learning and the use of gestures in bilingual education. Further studies in the field should disentangle the impact of gestures in L2 practice with regard to the different phases of learning: from encoding to training with or without consolidation phases in order to make L2 word acquisition more efficient and retention durable.

AUTHOR CONTRIBUTIONS

MM conceived the project and contributed to experimental design, stimulus material, behavioral and fMRI data collection, and paper writing. MM and CR organized the project. CR contributed to behavioral and fMRI data collection, behavioral data analysis, paper and writing in behavioral section. AI contributed to fMRI data analysis. KM contributed to critical discussion of data and manuscript. CR, MM, and AI created tables and figures and contributed to review and critique.

REFERENCES

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol.* 128, 466–478. doi: 10.1016/j.actpsy.2008.03.014

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., et al. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cereb. Cortex* 22, 2076–2086. doi: 10.1093/cercor/bhr287

Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguistics* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003

Aridan, N., and Mukamel, R. (2016). Activity in primary motor cortex during action observation covaries with subsequent behavioral changes in execution. *Brain Behav.* 6:e00550. doi: 10.1002/brb3.550

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *NeuroImage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Atkinson, R. C., and Shriffin, R. M. (1968). "Human memory: a proposed system and its control processes," in *The Psychology of Learning and Motivation*, Vol. 2, ed. K. W. Spence (New York, NY: Academic Press), 89–195.

Baddeley, A. D., and Andrade, J. (2000). Working memory and the vividness of imagery. *J. Exp. Psychol. Gen.* 129, 126–145. doi: 10.1037/0096-3445.129.1.126

Binder, J. R., Desai, R., Graves, W., and Conant, L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cognit. Sci.* 15, 527–536. doi: 10.1016/j.tics.2011.10.001

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage* 25, 958–968. doi: 10.1016/j.neuroimage.2004.12.019

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404. doi: 10.1046/j.1460-9568.2001.01385.x

Buccino, G., Binkofski, F., and Riggio, L. (2004a). The mirror neuron system and action recognition. *Brain Lang.* 89, 370–376. doi: 10.1016/S0093-934X(03)00356-0

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., et al. (2004b). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron* 42, 323–334. doi: 10.1016/S0896-6273(04)00181-3

Cattaneo, Z., Renzi, C., Casali, S., Silvanto, J., Vecchi, T., Papagno, C., et al. (2014). Cerebellar vermis plays a causal role in visual motion discrimination. *Cortex* 58, 272–280. doi: 10.1016/j.cortex.2014.01.012

Cohen, R. L. (1981). On the generality of some memory laws. *Scand. J. Psychol.* 22, 267–281. doi: 10.1111/j.1467-9450.1981.tb00402.x

Craik, F., and Lockhart, R. (1972). Levels of processing – Framework for memory research. *J. Verb. Learn. Verb. Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Craik, F. I., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104, 268–294. doi: 10.1037/0096-3445.104.3.268

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., and Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *J. Cogn. Neurosci.* 21, 803–820. doi: 10.1162/jocn.2009.21059

Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111

Dietz, N. A. E., Jones, K. M., Gareau, L., Zeffiro, T. A., and Eden, G. F. (2005). Phonological decoding involves left posterior fusiform gyrus. *Hum. Brain Mapp.* 26, 81–93. doi: 10.1002/hbm.20122

Ekuni, R., Vaz, L. J., and Bueno, O. F. A. (2011). Levels of processing: the evolution of a framework. *Psychol. Neurosci.* 4, 333–339. doi: 10.3922/j.psns.2011.3.006

Engelkamp, J., and Jahn, P. (2003). Lexical, conceptual and motor information in memory for action phrases: a multi-system account. *Acta Psychol.* 113, 147–165. doi: 10.1016/S0001-6918(03)00030-1

Engelkamp, J., and Krumnacker, H. (1980). Imaginale und motorische prozesse beim behalten verbalen materials. *Z. Exp. Angew. Psychol.* 27, 511–533.

Engelkamp, J., and Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *Germ. J. Psychol.* 9, 239–254.

Engelkamp, J., and Zimmer, H. D. (1994). *Human Memory: A Multimodal Approach*. Seattle: Hogrefe & Huber.

Engelkamp, J., Zimmer, H. D., Mohr, G., and Sellen, O. (1994). Memory of self-performed tasks: self-performing during recognition. *Mem. Cognit.* 22, 34–39. doi: 10.3758/BF03202759

Eschen, A., Freeman, J., Dietrich, T., Martin, M., Ellis, J., Martin, E., et al. (2007). Motor brain regions are involved in the encoding of delayed intentions: a fMRI study. *Int. J. Psychophysiol.* 64, 259–268. doi: 10.1016/j.ijpsycho.2006.09.005

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., and Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *NeuroImage* 10, 385–396. doi: 10.1006/nimg.1999.0484

Gatti, R., Rocca, M. A., Fumagalli, S., Cattrysse, E., Kerckhofs, E., Falini, A., et al. (2016). The effect of action observation/execution on mirror neuron system recruitment: an fMRI study in healthy individuals. *Brain Imaging Behav.* 11, 565–576. doi: 10.1007/s11682-016-9536-3

Grant, A. M., Fang, S. Y., and Li, P. (2015). Second language lexical development and cognitive control: a longitudinal fMRI study. *Brain Lang.* 144, 35–47. doi: 10.1016/j.bandl.2015.03.010

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Mapp.* 30, 3309–3324. doi: 10.1002/hbm.20753

Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., and Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends Cognit. Sci.* 19, 374–382. doi: 10.1016/j.tics.2015.05.006

Hodzic, A., Kaas, A., Muckli, L., Stirn, A., and Singer, W. (2009). Distinct cortical networks for the detection and identification of human body. *NeuroImage* 45, 1264–1271. doi: 10.1016/j.neuroimage.2009.01.027

Holle, H., Gunter, T. C., Rüschemeyer, S.-A., Hennenlotter, A., and Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage* 39, 2010–2024. doi: 10.1016/j.neuroimage.2007.10.055

Hunt, R. R., and Worthen, J. B. (2006). *Distinctiveness and Memory*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195169669.001.0001

Kelly, S. D., Hansen, B. C., and Clark, D. T. (2012). "Slight" of hand: the processing of visually degraded gestures with speech. *PLoS One* 7:e42620. doi: 10.1371/journal.pone.0042620

Kelly, S. D., McDevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Lang. Cognit. Process.* 24, 313–334. doi: 10.1080/01690960802365567

Kirchhoff, B. A., Wagner, A. D., and Maril, A. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *J. Neurosci.* 20, 6173–6180. doi: 10.1523/JNEUROSCI.20-16-06173.2000

Klimesch, W. (1987). A connectivity model for semantic processing. *Psychol. Res.* 49, 53–61. doi: 10.1007/bf00309203

Klimesch, W. (1994). *The Structure of Long-Term Memory: A Connectivity Model of Semantic Processing*. Hillsdale, NJ: Erlbaum.

Kormi-Nouri, R. (1995). The nature of memory for action events: an episodic integration view. *Eur. J. Cognit. Psychol.* 7, 337–363. doi: 10.1080/09541449508403103

Kormi-Nouri, R. (2000). The role of movement and object in action memory. *Int. J. Psychol.* 35, 132–132.

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

Levy, D. A., Bayley, P. J., and Squire, L. R. (2004). The anatomy of semantic knowledge: medial vs. lateral temporal lobe. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6710–6715. doi: 10.1073/pnas.0401679101

Liakakis, G., Nickel, J., and Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—A meta-analysis of neuroimaging studies. *Behav. Brain Res.* 225, 341–347. doi: 10.1016/j.bbr.2011.06.022

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

Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Front. Psychol.* 5:1467. doi: 10.3389/fpsyg.2014.01467

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

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

Macedonia, M., and Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Front. Psychol.* 7:953. doi: 10.3389/fpsyg.2016.00953

Macedonia, M., Müller, K., and Friederici, A. (2010). Neural correlates of high performance in foreign language vocabulary learning. *Mind Brain Educ.* 4, 125–134. doi: 10.1111/j.1751-228X.2010.01091.x

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

Macedonia, M., and Von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics* 6, 393–416.

Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., and Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain Res.* 1101, 102–109. doi: 10.1016/j.brainres.2006.05.033

Mayer, K. M.,. Yildiz, I. B., Macedonia, M., and von Kriegstein, K. (2015) Visual and motor cortices differentially support the translation of foreign language words. *Curr. Biol.* doi: 10.1016/j.cub.11.068

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cognit. Sci.* 7, 293–299. doi: 10.1016/S1364-6613(03)00134-7

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295X.102.3.419

Mulligan, N. W., and Hornstein, S. L. (2003). Memory for actions: self-performed tasks and the reenactment effect. *Mem. Cognit.* 31, 412–421. doi: 10.3758/BF03194399

Nilsson, L. G. (2000). "Remembering actions and words," in *The Oxford Handbook of Memory*, eds F. I. M. Craik and E. Tulving (Oxford: Oxford University Press), 137–148.

Nilsson, L. G., Nyberg, L., Klingberg, T., Aberg, C., Persson, J., and Roland, P. E. (2000). Activity in motor areas while remembering action events. *Neuroreport* 11, 2199–2201. doi: 10.1097/00001756-200007140-00027

Nyberg, L. (2002). Levels of processing: a view from functional brain imaging. *Memory* 10, 345–348. doi: 10.1080/09658210244000171

Ogawa, K., and Inui, T. (2011). Neural representation of observed actions in the parietal and premotor cortex. *NeuroImage* 56, 728–735. doi: 10.1016/j.neuroimage.2010.10.043

Oh, A., Duerden, E. G., and Pang, E. W. (2014). The role of the insula in speech and language processing. *Brain Lang.* 135, 96–103. doi: 10.1016/j.bandl.2014.06.003

Otten, L. J., Henson, R. N. A., and Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: relationship between findings from across- and within-task comparisons. *Brain* 124, 399–412. doi: 10.1093/brain/124.2.399

Ozyurek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130296. doi: 10.1098/rstb.2013.0296

Ozyurek, A., Willems, R. M., Kita, S., and Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cognit. Neurosci.* 19, 605–616. doi: 10.1162/jocn.2007.19.4.605

Paivio, A. (2006). Dual coding theory and education. *Paper presented at the Pathways to Literacy Achievement for High Poverty Children*, The University of Michigan School of Education, Ann Arbor.

Paulesu, E., Vallar, G., Berlingeri, M., Signorini, M., Vitali, P., Burani, C., et al. (2009). Supercalifragilisticexpialidocious: how the brain learns words never heard before. *NeuroImage* 45, 1368–1377. doi: 10.1016/j.neuroimage.2008.12.043

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Front. Neurosci.* 8:253. doi: 10.3389/fnins.2014.00253

Plass, J. L., Chun, D. M., Mayer, R. E., and Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *J. Educ. Psychol.* 90, 25–36. doi: 10.1037/0022-0663.90.1.25

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* 62, 816–847. doi: 10.1016/j.neuroimage.2012.04.062

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

Raboyeau, G., Marcotte, K., Adrover-Roig, D., and Ansaldo, A. I. (2010). Brain activation and lexical learning: the impact of learning phase and word type. *NeuroImage* 49, 2850–2861. doi: 10.1016/j.neuroimage.2009.10.007

Rizzolatti, G. (2015). "Action understanding," in *Brain Mapping*, ed. A. W. Toga (Waltham: Academic Press), 677–682. doi: 10.1016/B978-0-12-397025-1.00354-7

Rodriguez-Fornells, A., Cunillera, T., Mestres-Misse, A., and de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3711–3735. doi: 10.1098/rstb.2009.0130

Rolls, E. T. (2015). Limbic systems for emotion and for memory, but no single limbic system. *Cortex* 62, 119–157. doi: 10.1016/j.cortex.2013.12.005

Rueckl, J. R., Paz-Alonso, P. M., Molfese, P. J., Kuo, W., Bick, A., Frost, S. J., et al. (2015). Universal brain signature of proficient reading: evidence from four contrasting languages. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15510–15515. doi: 10.1073/pnas.1509321112

Sale, P., and Franceschini, M. (2012). Action observation and mirror neuron network: a tool for motor stroke rehabilitation. *Eur. J. Phys. Rehabil. Med.* 48, 313–318.

Saltz, E., and Donnenwerthnolan, S. (1981). Does motoric imagery facilitate memory for sentences – A selective interference test. *J. Verb. Learn. Verb. Behav.* 20, 322–332. doi: 10.1016/S0022-5371(81)90472-2

Segawa, J. A., Tourville, J. A., Beal, D. S., and Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *J. Cognit. Neurosci.* 27, 819–831. doi: 10.1162/jocn_a_00737

Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cognit. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006

Shams, L., Wozny, D. R., Kim, R. S., and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Front. Psychol.* 2:64. doi: 10.3389/fpsyg.2011.00264

Squire, L. R., Stark, C. E., and Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306. doi: 10.1146/annurev.neuro.27.070203.144130

Straube, B., Green, A., Weis, S., Chatterjee, A., and Kircher, T. (2008). Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *J. Cognit. Neurosci.* 21, 821–836. doi: 10.1162/jocn.2009.21053

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

Vogt, B. A., Finch, D. M., and Olson, C. R. (1992). Functional heterogeneity in cingulate cortex: the anterior executive and posterior evaluative regions. *Cereb. Cortex (New York, N.Y.: 1991)* 2, 435–443. doi: 10.1093/cercor/2.6.435-a

Wagner, A. D., Schacter, D. L., Rotte, M., and Koutstaal, W. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281, 1189–1191. doi: 10.1126/science.281.5380.1188

Wandell, B. A. (2011). The neurobiological basis of seeing words. *Ann. N. Y. Acad. Sci.* 1224, 63–80. doi: 10.1111/j.1749-6632.2010.05954.x

Whitney, C., Kirk, M., O'sullivan, J., and Ralph, M. (2010). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cereb. Contex* 21, 1066–1075. doi: 10.1093/cercor/bhq180

Willems, R. M., Özyürek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Ye, Z., Mestres-Misse, A., Rodriguez-Fornells, A., and Munte, T. F. (2011). Two distinct neural networks support the mapping of meaning to a novel word. *Hum. Brain Mapp.* 32, 1081–1090. doi: 10.1002/hbm. 21093

Zaccarella, E., and Friederici, A. D. (2015). Reflections of word processing in the insular cortex: a sub-regional parcellation based functional assessment. *Brain Lang.* 142, 1–7. doi: 10.1016/j.bandl.2014. 12.006

Zimmer, H. D. (2001). *Memory for Action: A Distinct form of Episodic Memory?* Oxford: Oxford University Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JC and handling Editor declared their shared affiliation.