Impact of multisensory learning on perceptual and lexical processing of unisensory Morse code

F.B. Junker [a,c,*], L. Schlaffke [b], N. Axmacher [a], T. Schmidt- Wilcke [c,d]

[a] Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr-University Bochum, Universitätsstraße 150, D-44801 Bochum, Germany
[b] Department for Neurology, BG-University Hospital Bergmannsheil, Bürkle de la Camp-Platz 1, D-44789 Bochum, Germany
[c] Department of Clinical Neuroscience and Medical Psychology, Heinrich Heine University, Universitätsstraße 1, D-40225 Düsseldorf, Germany
[d] Department of Neurology, St. Mauritius Clinic, Strümper Str. 111, D-40670 Meerbusch, Germany

ABSTRACT

Multisensory learning profits from stimulus congruency at different levels of processing. In the current study, we sought to investigate whether multisensory learning can potentially be based on high-level feature congruency (same meaning) without perceptual congruency (same time) and how this relates to changes in brain function and behaviour. 50 subjects learned to decode Morse code (MC) either in unisensory or different multisensory manners. During unisensory learning, the MC was trained as sequences of auditory trains. For low-level congruent (perceptual) multisensory learning, MC was applied as tactile stimulation to the left hand simultaneously to the auditory stimulation. In contrast, high-level congruent multisensory learning involved auditory training, followed by the production of MC sequences requiring motor actions and thereby excludes perceptual congruency. After learning, group differences were observed within three distinct brain regions while processing unisensory (auditory) MC. Both types of multisensory learning were associated with increased activation in the right inferior frontal gyrus. Multisensory low-level learning elicited additional activation in the somatosensory cortex, while multisensory high-level learners showed a reduced activation in the inferior parietal lobule, which is relevant for decoding MC. Furthermore, differences in brain function associated with multisensory learning was related to behavioural reaction times for both multisensory learning groups. Overall, our data support the idea that multisensory learning is potentially based on high-level features without perceptual congruency. Furthermore, learning of multisensory associations involves neural representations of stimulus features involved in learning, but also share common brain activation (i.e. the right IFG), which seems to serve as a site of multisensory integration.

1. Introduction

Although human experience is multisensory in nature, previous research, especially in neuroimaging studies, has predominantly focused on learning and memory formation in the context of unisensory stimulus processing (Alais et al., 2010). The use of more than one modality provides both redundant and complementary information and has frequently been shown to be associated with superior learning effects (for a review, see Matusz et al., 2017). Specifically, in language comprehension multisensory (MS) processing plays a critical role. While language is learned during childhood primarily via (spoken) speech, other sensory inputs such as facial expressions, gestures and lip movements additionally shape the processing of language (Rowe and Goldin-Meadow, 2009; Vigliocco et al., 2014; Özyürek, 2014). This additional information is supposed to strongly support the formation of MS feature representations like semantics (Doehrmann and Naumer, 2008) that, once built, can be reactivated by a variety of sensory inputs (Yildirim and Jacobs, 2013). Hence, language is a MS phenomenon involving high-level feature representations independent of the sensory modality (Fatma et al., 2019).

Current research indicates that MS integration is stronger, if two stimuli [1] arise at the same time (temporal rule), [2] arise from the same location (spatial rule) and [3] are relatively weak (inverse effectiveness rule) (Lakatos et al., 2007; Alais et al., 2010; Klemen and Chambers, 2012). In addition to these low-level feature congruency (same time and location), MS integration can also be affected by high-level features like semantics (Doehrmann and Naumer, 2008; Calvert and Thesen, 2004). In the so called McGurk effect, the integration of auditory-visual information is manipulated by incongruent MS information leading to an illusion (for details, see McGurk and MacDonalds, 1976) specially in persons with higher experiences in speech processing (adults > childs; Choi et al., 2018).

Successful integration is key to MS learning. Importantly it could be demonstrated that MS learning influences subsequent unisensory processing (Lehmann and Murray, 2005; Seitz et al., 2006; Shams and Seitz, 2008; Shams et al., 2011), i.e. at a neural level exposure of congruent auditory-visual stimuli during learning changes the way auditory and visual stimuli are processed when presented separately (McIntosh et al., 1998; Nyberg et al., 2000). Three mechanisms [M1-3] for MS learning (Fig. 1: Induced plasticity) are currently being discussed and supported by human research (for a review, see Shams and Seitz, 2008; Paraskevopoulos and Herholz, 2013). [M1] Synchronized MS stimulation can induce plasticity in modality-specific feature representations (Lappe et al., 2008, 2011; Chen et al., 2012), possibly involving Hebbian learning (Seitz and Dinse, 2007; see Fig. 1). [M2] MS learning can build or change structural (Scholz et al., 2009) and functional (von Kriegstein and Giraud, 2006) connectivity between unisensory feature representations, or [M3] MS learning can build or change MS feature representations or their connections to unisensory representations (Naumer et al., 2009; Paraskevopoulos et al., 2012). All three mechanisms can explain enhanced behaviour after MS learning, while each mechanism leads to different predictions of the neuronal responses to unisensory stimuli (Fig. 1: Unisensory response). [M1] Plasticity in unisensory feature representations leads to changed response of these representations. [M2] Plasticity in unisensory interconnections leads to a coactivation of both unisensory representations. Or [M3] plasticity in MS representations or their connections leads to changed response within these MS representations.

While the representation and integration of low-level features (e.g. time) is related to lower stages of the cortical processing hierarchy (e.g. auditory-visual integration in V5 and *planum temporale;* Saldern and Noppeney, 2013), high-level features (e.g. lexical, semantic) affect processing at higher stages (e.g. auditory-visual integration in *inferior parietal sulcus;* Sadaghiani et al., 2009). Current research indicates that MS features are represented as associations at different levels of processing (Shamsand& Seitz, 2008; Werner and Noppeney, 2010; Xi et al., 2019), where more abstract features (high-level > low-level) are processed at higher cortical areas in the processing hierarchy (Calvert and Thesen, 2004). In contrast to low-level feature integration, where spatial and temporal congruency is necessary for a successful integration at early stages (within 100 ms post-stimulus), high-level features are integrated at later stages (>100 ms post-stimulus; ten Oever et al., 2016) involving top-down modulation (Choi et al., 2018). So far, it has not been investigated, [Q1] whether MS learning can be mediated without low-level congruency (different time), but instead using high-level MS features (same meaning; Fig. 1: M3). And if so, [Q2] whether MS learning involving low- and high-level is mediated via plasticity involving different feature representations (unisensory vs. multisensory representations), as expected by mechanisms of MS learning (Fig. 1: M1/2 vs. M3).

To answer these questions, we analysed and compared behaviour; i. e. reaction times and decoding accuracy (performance) and related neural activation (fMRI) during perceptual (task 1) and lexical processing (task 2) of auditory Morse code in three groups that had undergone different learning strategies during MC training. Both tasks used the same auditory stimuli (combinations of Morse code letters) while the subjects either had to compare the length of the first and last tone (same or different), or had to make a lexical decision (word or nonword).

Morse code (MC) is a method for encoding language as combinations of short and long stimuli (e.g. sounds, vibrations, lights) and can be processed at different feature levels (Schlaffke et al., 2015), ranging from perceptual (low-level) to lexical and semantic features (high-level). MC per se is amodal, meaning that the same set of stimuli can be applied using different sensory modalities (e.g. auditory training: Group 1 - US) without changing their features. By processing MC across multiple sensory modalities, associations can be learned based on different
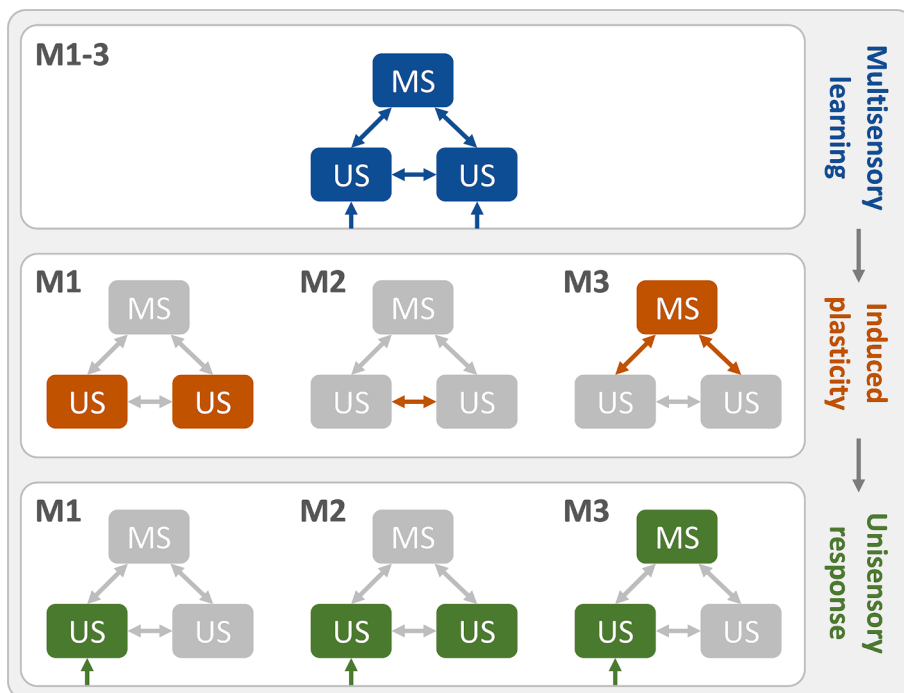
Fig. 1. Mechanism for multisensory learning. Schematic representation of enhanced unisensory processing after multisensory (MS) learning. MS stimulation (first row) can induce plasticity (second row) leading to different predictions of neuronal responses during unisensory processing (third row). MS learning can induce plasticity [M1] in unisensory representation (US) or [M2] their interconnections, as well as in [M3] MS representations and their connections to unisensory representation. Of note, the single arrows (bottom of third row) indicate unisensory input after unisensory or multisensory learning as performed in our experiment.

feature levels (levels of associations). During synchronized MS stimulation (auditory-tactile), the MC can be integrated and associated involving low-level features (same time: Group 2 – MS-low). In contrast, a training separated in time (auditory/motor) violates low-level congruency. Instead, associations can only be learned involving high-level features like the lexical meaning of individual MC letters (Group 3 – MS-high; see Table 1). If MS learning involving only high-level features (same lexical meaning of MC letters) is possible [Q1], we expect to find differences in behaviour and/or neural activation compared to unisensory learners (group 3 vs group 1). As predicted by mechanism 3 (Fig. 1: M3), these changes should be related to high-level representations relevant for translating individual MC letters [Q2] and thereby should not facilitate perceptual processing, where these representations are not involved (analysis 2). In contrast, MS associations at low-level features (same time) should lead to an improvement in perceptual processing (analysis 1) as compared to unisensory training. Furthermore, the neural origin associated with different MS learning strategies gives further indication about the mechanism involved in MS learning (group 2 vs group 3).

2. Results

2.1. Behavioural data

To evaluate the behavioural effect of different learning strategies (unisensory, MS-low, MS-high), recognition performance and reaction time during perceptual (task 1: light colour) and lexical processing (task 2: dark colour) were analysed and compared. No difference in performance was found between the three groups, neither for the perceptual task ($p_{Kruskal-Wallis} = 0.065$) nor for the lexical-decision task ($p_{Kruskal-Wallis} = 0.184$). The subjects identified significantly more stimuli in the perceptual task as compared to the lexical-decision task (see Fig. 2a) in all groups (US: $p_{Wilcoxon} = 0.005$; MS-low: $p_{Wilcoxon} = 0.016$; MS-high: $p_{Wilcoxon} = 0.004$). Differences in reaction times (see Fig. 2b) were found for the perceptual task ($p_{Kruskal-Wallis} = 0.036$), where subjects showed faster reaction times after MS-low learning compared to unisensory learning ($p_{Kruskal-Wallis} = 0.033$). Not surprisingly, more time was needed for lexical compared to perceptual processing in all groups (US: $p_{t-test} < 0.001$; MS-low: $p_{t-test} < 0.001$; MS-high: $p_{Wilcoxon} < 0.001$). For further information regarding data distributions, see Table 2.

2.2. fMRI data

fMRI data were analysed as a 3×2 analysis of variance (ANOVA; group and task as factors). As we were primarily interested in the activation pattern related to different learning strategies (Q1, Q2), activation related to the different tasks (Main effect: Task) was not investigated in the current analyses. These effects were already discussed by Schlaffke and colleagues (2015) who could show that perceptual processing stronger relies on the superior temporal gyri as well as the supplementary motor area bilaterally. In contrast, lexical processing showed enhanced activation of the left inferior frontal gyrus and the left occipitotemporal cortex.

Table 1

Learning. Learning and testing procedure for all groups with corresponding group-size (n). In unisensory training (US), MC was learned using only auditory stimuli (A). In addition, multisensory training involved either a synchronized vibrotactile stimulation (VT) of the left hand (MS-low), or a separate motor (M) training via active MC tapping (MS-high).

Group		n	Learning	Test
US:	Unisensory	17	A	A
MS-low:	Multisensory low-level	16	A + VT	A
MS-high:	Multisensory high-level	17	A/M	A

A: Auditory VT: Vibrotactile M: Motor.

2.2.1. Analysis 1 – Main effect: Learning

To identify brain regions more involved in perceptual and lexical processing of MC dependent on the learning strategy, the main effect 'learning' was analysed. Three brain regions showing different activation depending on the learning strategies (US, MS-low, MS-high) were identified ($p_{ANOVA} < 0.001$; Fig. 3a/Table 3.1). Post-hoc tests (Table 3.2) revealed that the activation in the right *inferior frontal gyrus* (136 voxels, $z = 4.52$) was significantly increased in both MS learning groups as compared to the unisensory group (Fig. 3b-2). MS low-level learners showed additional activation in the *midcingulate cortex* (139 voxels, $z = 4.43$) extending into the right *postcentral gyrus* (Fig. 3b-3) as compared to the two other groups. Reduced activity in left *supramarginal gyrus* (270 voxels; $z = 5.04$) was found after MS high-level learning as compared to unisensory and MS-low learners (see Fig. 3b-1).

To test for the relationship of brain activations and behaviour (performance, reaction time) multiple linear regression analyses were performed for each learning group separately. While brain activations were not able to predict decoding performance (US: $p = .592$; MS-low: $p = .224$; MS-high: $p = .553$; see Table 4.1), a relationship to reaction times was found for both MS learning groups (US: $p = .237$; MS-low: $p = .016$; MS-high: $p = .02$; see Table 4.2). Activation within the right IFG (Beta = 0.47, $p = .046$) and *midcingulate cortex* (Beta = -0.686, $p = .002$) predicted reaction times after low-level MS learning. For high-level MS learners, reaction times after learning were related to IFG (Beta = -0.427, $p = .012$) and IPL (Beta = -0.348, $p = .04$) activation.

2.2.2. Analysis 2 – Interaction: Learning × Task

To identify brain regions more involved in perceptual than lexical processing of MC dependent on the learning strategy, the interaction effect of different learning strategy and performed task was analysed. No brain regions showed activation dependent of the learning strategy (US, MS-low, MS-high) and the performed task (perceptual, lexical-decision). Brain activations were either related to the task (not discussed here) or the learning strategy (Analysis 1, see above) and did not interact.

3. Discussion

Using MC as a model for learning, we investigated how unisensory and MS learning strategy affect perceptual processing as well as lexical-decision making. At the behavioural level, superior learning effects were found during perceptual processing where multisensory low-level learning (MS-low) resulted in faster reaction times compared to unisensory learning. At the neural level, activity in the *inferior frontal gyrus* (IFG), *inferior parietal lobule* (IPL) and *cingulate cortex/postcentral gyrus* was independent of the task, but related to the learning strategy (Analysis 1 – Main effect: Learning). While MS learning (low- and high-level) led to additional activity at the right *inferior frontal gyrus*, only MS low-level learners showed increased activity in the right *cingulate cortex/postcentral gyrus*. In contrast, MS-high learners showed reduced activity at the left IPL (*supramarginal gyrus*) as a consequence of auditory and motor training separated in time. Furthermore, behavioural reaction times were related to the degree in neural activations within regions that were modulated by MS low- and high-level learning strategies, respectively.

MS low-level learners, where MC had been presented synchronously as auditory and vibrotactile stimulation (during learning), showed faster reaction times during perceptual processing that was related to changes in brain function occurred during learning. In contrast, MS-high learners had a temporal gap between auditory stimulation and MC production, i.e. with regard to MS principles, integration of low-level feature is not likely to have occurred (temporal rule; Calvert, 2001; Calvert and Thesen, 2004). Importantly, a beneficial effect could not be found for higher lexical-decision making after MS-low learning. This enhanced perceptual processing is not likely to facilitate high-level lexical-decision making, which includes additional processing steps. While a beneficial effect of MS-high learning on behaviour was not found,
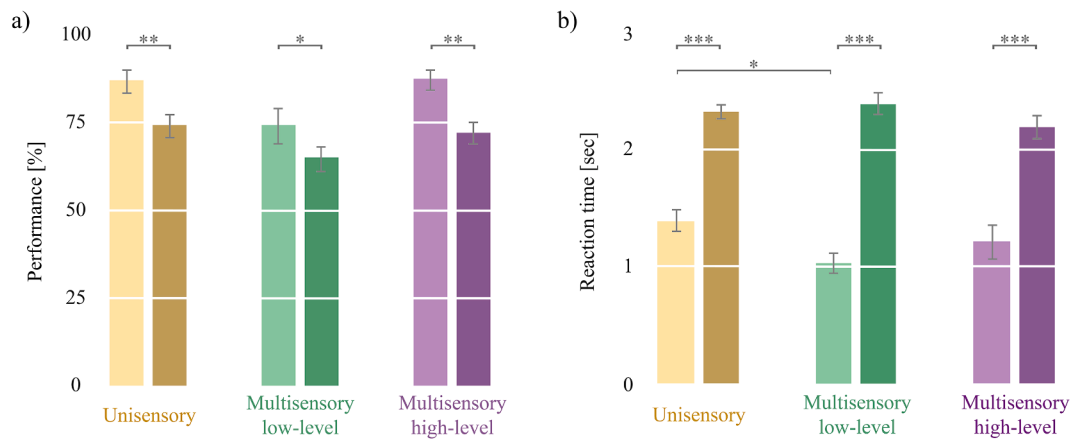
Fig. 2. Performance and reaction times. Behavioural results (a: performance; b: reaction time) of the perceptual (light) and lexical-decision task (dark) for unisensory (yellow), MS low-level learners (green) and MS high-level (violet). Significant differences are marked (∗: p < .05; ∗∗: p < .01; ∗∗∗: p < .001).

Table 2

Behavioural distribution. Statistical test for normality (Shapiro-Wilk) of behavioural data. Performances and reaction times were analysed separately for each group and task (P = perceptual; LD = lexical-decision).

Group	Performance		Reaction time	
	P	LD	P	LD
Unisensory	0.002	0.962	0.998	0.636
Low-level multisensory	0.028	0.475	0.426	0.412
High-level multisensory	0.002	0.033	0.019	0.729

neither for lower perceptual processing nor higher lexical-decision making, a relationship between brain activation and reaction time similar to synchronized MS training shows that the learning strategy is able to modify brain function and possibly thereby enhance behaviour. However, a facilitating effect on reaction time was not found here, possibly due to the rather short period of motor-training (5 min per session) that prevented an impact on performance.

Increased activity (as compared to unisensory learners) was observed during perceptual processing as well as lexical-decision making of auditory MC in the right *inferior frontal gyrus* (IFG) after both, MS-low and MS-high learning. An involvement of the right IFG in MS processing has previously been described for various modalities (Hein et al., 2007) and different feature levels (e.g. spatial features: Ehrsson et al., 2005; lexical features: Noppeney et al., 2008). In the so-called "rubber-hand-illusion" (Botvinick and Cohen, 1998), an artificial hand (visible) is simultaneously touched with the natural hand (not visible), leading to the percept of body ownership driven by low-level features (same time and location). Ehrsson and colleagues (2005) could show that a successful fusion of visual and tactile information is related to activation in right *ventral premotor cortex* including the *inferior frontal gyrus*. The authors conclude that activation in the right IFG reflects MS integration. During auditory-visual integration of language, Van Atteveldt et al. (2007) found enhanced activation in the right IFG, specifically when incongruent combinations of graphemes (visual letters) and phonemes (sounds) were presented, suggesting right IFG involvement in auditory-visual integration of high-level stimulus features (e.g. lexical). Activation of the right IFG was also found in several studies using congruent (e. g. dog with barking sound) as well as incongruent semantic features (e.g. cat with barking sound). Enhanced activation of the right IFG for incongruent combinations was found, both for auditory-visual (Hein et al., 2007; Belardinelli et al., 2004; Noppeney et al., 2008) and auditory-tactile combinations (Hein et al., 2007), indicating that the right IFG is involved in the integration of high-level stimulus features (e. g. semantic) independent of the sensory modality (e.g. auditory, visual, tactile). Overall, these results suggest that the right IFG is involved in MS

integration at different levels of association (perceptual, lexical and semantic) regardless of the involved sensory modalities (auditory, visual, tactile and motor) or MS learning types (low-level, high-level). In contrast to feature-specific regions, the right IFG has a superior function in MS processing, by detecting violations/incongruency of learned associations (MC with motion/vibration; dog with barking) across sensory modalities (probably deciding which modality provides the correct information). Changes within the IFG after MS-high learning similar to MS-low learning therefore shows that MS learning using exclusively high-level features is possible without low-level (temporal) congruency [Q1].

During processing of auditory MC, MS-low learners showed additional activation in the right somatosensory cortex (*cingulate cortex/ postcentral gyrus*) contralateral to the site of tactile stimulation during training, suggesting that this region had become part of the processing network. This is an effect which has previously been described across multiple sensory modalities. After learning to associate an auditory stimulus with a visual event, McIntosh and colleagues (1998) found occipital activation to auditory stimulation without visual stimulation. Vice versa, by learning to associate a visual word with a sound, reactivation of auditory regions can be observed during unisensory visual word recognition only for those stimuli whose associations had been learned during training (Nyberg et al., 2000). Similar coactivations of the auditory cortex can be found during silent lipreading (Calvert et al., 1997). By presenting lip movement without the corresponding speech (visual only), reactivations of the auditory cortex were found, similar to those observed during processing of auditory speech. These reactivations were found only for lip movements related to language, where auditory-visual association were leaned throughout daily-life communication. After learning to associate specific voices with faces, von Kriegstein and Giraud (2006) found increased functional coupling between unisensory representations for voices (temporal voice area) and faces (fusiform face area), indicating an involvement of visual areas (fusiform face area) in auditory (only) voice recognition. The learning of auditory-tactile associations involving low-level perceptual congruency leads to changes in neural activation pattern during auditory (unisensory) MC processing, involving unimodal cortices (*postcentral gyrus*) not engaged in processing before training. In this regard, the result reflects learned association involving stronger functional coupling of the unisensory cortices as predicted for mechanism 2 (Fig. 1: M2).

MS-high learners showed reduced activity in the left *IPL* (*supramarginal gyrus*) which has been frequently reported during the processing of spoken and written language. In a *meta*-analysis based on 36 neuroimaging studies related to reading, Taylor and colleagues (2013) concluded, that the left IPL is specifically involved in the transformation of graphemes (written symbols) into phonemes (sounds) while reading.
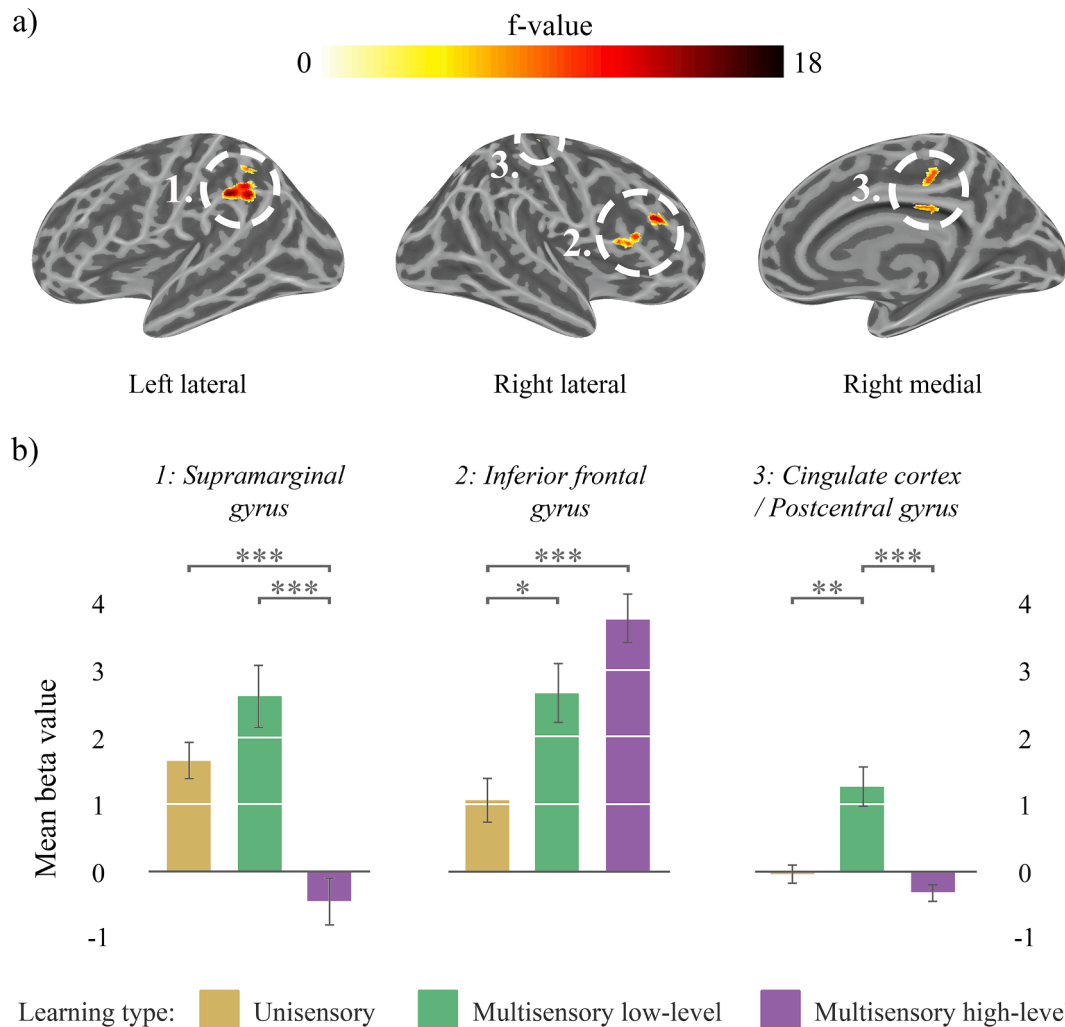
Fig. 3. Main effect: Learning. (a) Statistical parametric f-map of cortical brain activation family-wise error corrected (FWE) for multiple comparisons at the cluster level (p_{FWE} < 0.05) for the main effect of learning strategies (Analysis 1). (b) Corresponding beta values in each cluster for unisensory (yellow), multisensory low-level (green) and multisensory high-level learners (violet). Significant differences are marked (*: p < 0.05, **: p < 0.01, ***: p < .001; for details, see Table 3.2).

Table 3.1

Main effect "learning". Peak-cluster activations for the main effect 'learning' with corresponding MNI-coordinates (x,y,z), cluster size and z-values.

Cluster	x	y	z	z-Value	# Voxels
Inferior parietal lobule (Supramarginal)	−56	−38	44	5.04	270
Inferior frontal gyrus (Opercularis)	46	36	26	4.52	136
Cingulate cortex (Middle)	24	−24	44	5	139

Table 3.2

Post-hoc "learning". Post-hoc test for the main effect 'learning' with corresponding p-values for the comparison between unisensory (US), multisensory low-level learners (MS-low) and multisensory high-level (MS-high). According to the variance homogeneity post-hoc tests were performed using Bonferroni or Games-Howell test (marked with *).

Cluster	US vs. MS-low	US vs. MS-high	MS-low vs. MS-high
Inferior parietal lobule (Supramarginal)	0.197	<0.001	<0.001
Inferior frontal gyrus (Opercularis)	0.011	<0.001	0.115
Cingulate cortex (Middle)	0.001*	0.269*	<0.001*

Similar, Cattinelli et al. (2013) found two separate clusters in the left IPL after reanalysing 35 studies related to reading. While reading words is related to activation of the *angular gyrus*, pseudoword reading relies on grapheme-to-phoneme conversions within the *supramarginal gyrus*. These results are consistent with other studies (e.g. Protopapas et al., 2016; DeMarco et al., 2017) indicating a specific role of the left *supramarginal gyrus* in the conversion of encoded language (visual graphemes) into phonemes. Analogous to written language, decoding MC involves conversion of auditory sequences (MC) into phonemes (high-level feature). After MC training, increased activation of the left IPL can be observed during lexical-decision making as compared to before learning (Schlaffke et al., 2015), indicating a high relevance of the left IPL also in the lexical conversion of MC. The reduced activation within the IPL found here also fits to the observation that brain activations correlate negatively with beneficial MS learning (e.g. Matusz et al., 2015). By pairing visual objects with meaningless sounds, typically associated with decreased learning performance (Matusz et al., 2017), Thelen and colleagues found enhanced ERPs within high-level representations (*Middle temporal gyrus*) that were linked to reduced memory performance (Thelen et al., 2012). With respect to MS-high learners, where MS associations can only be learned based on their lexical feature (high level multisensory), reduced *supramarginal/IPL* activity might increase efficiency. Although a facilitating effect on behaviour was not found here, a more efficient neural processing indicates that behaviour can be

Table 4.1
Main effect 'learning' on behavioural performance. Multiple linear regressions for testing the effect of brain activations (Main effect: 'learning') on behavioural performance for each learning group. The p-value per model (p $_{model}$), as well as the p-values (p $_{var}$) and standardized correlation coefficient (Beta) for each brain region is given.

Cluster	Performance								
	Unisensory			MS-low			MS-high		
	p $_{model}$	p $_{var}$	Beta	p $_{model}$	p $_{var}$	Beta	p $_{model}$	p $_{var}$	Beta
Inferior parietal lobule (supramarginal)	0.592	0.559	0.111	0.244	0.270	0.264	0.553	0.744	0.061
Inferior frontal gyrus (opercularis)		0.562	0.110		0.842	0.050		0.943	-0.013
Cingulate cortex (middle)		0.440	0.141		0.651	0.104		0.170	-0.264

Table 4.2
Main effect 'learning' on behavioural reaction time. Multiple linear regressions for testing the effect of brain activations (Main effect: 'learning') on behavioural reaction time for each learning group. The p-value per model (p $_{model}$), as well as the p-values (p $_{var}$) and standardized correlation coefficient (Beta) for each brain region is given.

Cluster	Reaction time								
	Unisensory			MS-low			MS-high		
	p $_{model}$	p $_{var}$	Beta	p $_{model}$	p $_{var}$	Beta	p $_{model}$	p $_{var}$	Beta
Inferior parietal lobule (supramarginal)	0.237	0.563	0.105	0.016	0.380	0.188	0.020	0.040	−0.348
Inferior frontal gyrus (opercularis)		0.065	−0.345		0.046	0.470		0.012	−0.427
Cingulate cortex (middle)		0.281	0.190		0.002	−0.683		0.050	0.337

enhanced after longer training (only 30 min here). The results not only indicate that multi-sensory learning based on high-level features (lexical meaning) is possible [Q1], it also shows that this learning process is based on high-level feature representations not involved during synchronized (low-level feature) MS training, as expected by mechanism 3 [Q2].

3.1. Conclusion

By using MC, we were able to investigate the impact of multisensory learning on unisensory processing, involving different feature-levels. Here, our data suggest that associations across modalities can be learned without low-level perceptual congruency using exclusively high-level features (e.g. lexical).

While associations of low-level features were associated with increased functional coupling of unisensory representations (here: auditory-somatosensory) and faster reaction times during perceptual processing, the learning of high-level feature associations involved changes within feature-specific representations (here: *inferior parietal lobule*). Additionally, activation of the right *inferior frontal gyrus* (IFG) was found for both multisensory learning types, suggesting an important role of the right IFG in multisensory integration (Van Atteveldt et al., 2007; Belardinelli et al., 2004; Noppeney et al., 2008; Ehrsson et al., 2005) and learning (Hein et al., 2007), independent of the modalities or feature-levels.

Our data suggest that neural activation during processing is not only related to the task per se, i.e. stimulus material and modality, but also to the individual experience, i.e. the learning strategy.

4. Experimental procedure

4.1. Morse code

The international Morse code (MC) is a method for text encoding by using standardized sequences of short ("dots" [●]) and long ("dashes" [—]) signals. As series of on and off sounds, vibrations or lights, MC encodes the Latin letters, Arabic numbers and some basic punctuation. Therefore, MC can be used as a model for language learning. The meaning of a stimulus depends on its sequence, not on the way it is presented; as such, the code is amodal.

4.2. Subjects

50 healthy, right-handed subjects (mean age 24.7 years, standard deviation = 2.9; 21 females) participated in the study. All participants were naive to MC prior to the learning intervention. 17 subjects learned to decipher auditory (unisensory) MC signals while passively listening (Group 1: Schlaffke et al., 2015). 16 subjects learned to decode MC, while the pattern was synchronously applied as auditory and vibrotactile sequences (Group 2). Another 17 participants learned to actively applying MC (5 min per session) separately to the auditory training (Group 3: Schlaffke et al., 2017). To participate in the study, subjects provided written consent. The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Ruhr-University Bochum, Germany.

4.3. Learning

Only a subset of 12 MC letters was learned using an audiobook and headphones (40 mm speaker, 20–20,000 Hz frequency range, 98 dB sensitivity; stimulation frequency: 802 Hz). On six days, study participants learned the letters in a specific order as a standardized procedure (day 1: E, S, N, and O; day 2: T and R; day 3: U and D; day 4: A and I; day 5: M and G; day 6: repetition of all letters). A repetition of the previously learned MC letters was performed at the beginning of all training days (apart from day 1). To complete each training session, the subjects had to reach a specific success criterion. If this criterion was not reached, the lesson was expanded by including additional MC letters (~5 min). Afterwards, all subjects finished the training sessions successfully. Unisensory learners (US; Group 1) only trained to decode auditory MC letters as described above. In addition, MC stimuli were synchronously applied as auditory and vibrotactile stimuli during multisensory low-level learning (MS-low; Group 2). For this purpose, the index and middle fingers of the left hand were stimulated by a piezo element (Dancer Design Mini-PTS stimulators, MPTS-008 control unit; Stimulation-frequency: 200 Hz). In contrast, high-level multisensory learners (MS-high; Group 3) practiced the transmittance of 60 MC letters with their right index finger on a touch-screen instead of decoding 60 auditory letters, as unisensory learners did (see Table 1). All groups followed a very similar learning protocol, spending the same amount of time on training (6x 25–30 min).

4.4. Task

fMRI recordings were conducted both, before and after MC learning. In trains of three letters MC was presented only acoustically, comprised 40 words (mean stimulus length: 3.57 s), 40 nonwords (mean stimulus duration: 3.56 s) and 25 times the SOS signal (stimulus duration: 2.28 s). In addition, 25 control tones (beep tone; 786 Hz, duration: 3 s) were presented, randomly ordered by using the stimulus delivery and experiment control software Presentation (Neurobehavioral Systems, USA). fMRI data before learning were not analysed in this study.

Participants were required to perform either a perceptual task (P) or a lexical-decision task (LD). In the perceptual task, subjects had to compare the length of the first and last stimuli against each other (same or different). In contrast, MC trains had to be decoded and labelled as words, nonwords, SOS or a control tone in the lexical-decision task. Subjects responded via button press with the left hand (word/nonword/ SOS/control tone). Since groups only differed with respect to the learning strategy (in how they learned the individual MC letters during training), the later analysis just involved those stimuli that either represented a word or nonword. For this reason, brain activation during word and nonword processing after learning were analysed together (allwords). Of note, brain activity related to SOS and control tone decoding was not subjected to the current analysis. Both stimuli were needed to analyse the data in the context of language decoding, similar to reading. The control tone was used to contrast the data for auditory stimulation and to analyse the processing of words and non-words in the context of non-lexical (sub-word) language decoding (see Junker et al., 2020). We are using the SOS as a model for learning and stimulus decoding in the context of fast lexical (whole-word) language processing.

4.5. fMRI sequences

Magnetic Resonance Imaging was performed on a 3.0 Tesla scanner (Philips Achieva 3.2, Best, Netherlands). Using a 32-channel head coil, high-resolution T1-weighted data sets (TR 8.3 ms, TE 3.8 ms, FOV 256 $\times$ 256, voxel size $1.00 \times 1.00 \times 1.00$ mm^3 reconstructed to $0.94 \times 0.94 \times 1.00$ mm^3) were acquired first. During T2*-weighted echo planar imaging (single shot EPI, 90° flip angle, TR 2400 ms, TE 35 ms, FOV 224×224 mm^2, voxel size $2 \times 2 \times 3$ mm^3, 36 slices, ascending scan order without gaps, 492 dynamic scans), the perceptual and lexical-decision task was performed.

4.6. Preprocessing and first level analyses

Using MRIconvert 2.0 (Lewis Center for Neuroimaging, University of Oregon, USA), functional images were converted from DICOM to NIfTI (HDR-IMG pairs) format. Preprocessing of functional images included slice-time correction, unwarping, realignment for movement correction, co-registration to the structural T1-image, spatial normalization to the same stereotactic space (using the SPM EPI-template) and spatial smoothing (full width at half maximum: 6 mm). Data processing was performed using SPM12 (Welcome Department of Cognitive Neurology, University College, London, UK) running under Matlab R2017a.

In the statistical framework of general linear models, one regressor ranging from stimulus onset to offset was created each for processing words, nonwords and the control tone and was convolved with the haemodynamic response function. Furthermore, the six movement parameters (three rotation parameters, three translation parameters) were added as covariates of no interest. Overall, first level analyses yielded two contrast (P_allwords, LD_allwords; P = perceptual analysis, LD = lexical-decision) per group (US, MS-low, MS-high).

4.7. Second level analyses

Second level analyses were performed as a 3x2 analysis of variance (ANOVA) within a flexible factorial design (task and learning strategy as factors with age and performance as covariates) involving only correctly identified stimuli. The main effects (if no interaction existed) were investigated for learning types only. As we were primarily interested in the activation patterns related to different learning histories, i.e. group differences, activations related to the different tasks, i.e. perceptual vs lexical-decision (levels of processing), were not subject to the current analyses (see Schlaffke et al., 2015). With an initial significance level of $p < .001$ corrected for multiple comparisons (family-wise error correction: $p < .05$ at the cluster level), voxel-wise whole brain analyses were performed. To identify effect directions, post-hoc t-tests were performed based on extracted mean beta-values for each cluster per subject and task. Labelling and visualisation was performed using the SPM12 extensions Automated Anatomical Labelling (http://www.gin.cnrs.fr/en/t ools/aal-aal2/) and bspmview (http://www.bobspunt.com/bspmview/).

4.7.1. Analysis 1 – Main effect: Learning

To identify brain regions more engaged after different learning strategies (US, MS-low, MS-high) independent of the task, the main effect of learning is analysed.

4.7.2. Analysis 2 – Interaction: Learning × Task

To identify brain regions more involved in the perceptual (P) or lexical-decision (LD) task dependent on the learning strategy (US, MS-low, MS-high), the interaction between task and learning strategy is analysed.

4.8. Behavioural relations

To further investigate if neural activations within regions that were modulated during different learning strategies (Analysis 1/2) can be related to behaviour, multiple linear regression analysis will be applied. Here, recognition performance and reaction time were tried to be predicted by the activation of these regions (identified in analysis 1 and 2) for each learning group.

5. Limitations

During MS-high learning, MC letters were transmitted by touching a screen with the right index finger. Because of this procedure, the motor training was accompanied by synchronized sensory feedback. Although the neural activation pattern does not suggest an involvement of the sensory system in decoding auditory MC after MS-high learning, an involvement of the sensory system during decoding cannot be fully excluded.

Another limitation that has to be mentioned is the limited amount of motor training during MS-high learning. The motor-training period was kept short, because the relevant MC-letter-associations had to be learned first, before a crossmodal transfer was possible. To keep the amount of training constant over all groups, the motor training was limited to 5 min.

CRediT authorship contribution statement

F.B. Junker: Methodology, Investigation, Formal analysis, Writing - original draft, Visualization. **L. Schlaffke:** Methodology, Investigation. **N. Axmacher:** Validation, Resources. **T. Schmidt- Wilcke:** Conceptualization, Validation, Resources, Writing - review & editing.

Germany for continuous scientific support.

References

Alais, D., Newell, F.N., Mamassian, P., 2010. Multisensory processing in review. physiology to behaviour. Seeing Perceiving 23 (1), 3–38. https://doi.org/10.1163/187847510X488603.

Van Atteveldt, N.M., Formisano, E., Goebel, R., Blomert, L., 2007. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. NeuroImage 36 (4), 1345–1360. https://doi.org/10.1016/j.neuroimage.2007.03.065.

Belardinelli, O.M., Sestieri, C., Di Matteo, R., Delogu, F., Del Gratta, C., Ferretti, A., Caulo, M., Tartaro, A., Romani, G.L., 2004. Audio-visual crossmodal interactions in environmental perception. An fMRI investigation. Cogn Process. 5 (3), 167–174. https://doi.org/10.1007/s10339-004-0024-0.

Botvinick, M., Cohen, J., 1998. Rubber hands 'feel' touch that eyes see. Nature. 391 (6669), 756. https://doi.org/10.1038/35784.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P. K., et al., 1997. Activation of auditory cortex during silent lipreading. Science (New York, N.Y.) 276 (5312), 593–596. https://doi.org/10.1126/science.276.5312.593.

Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11 (12), 1110–1123. https://doi.org/10.1093/cercor/11.12.1110.

Calvert, G.A., Thesen, T., 2004. Multisensory integration: methodological approaches and emerging principles in the human brain. J. Physiol. 98 (1–3), 191–205. https://doi.org/10.1016/j.jphysparis.2004.03.018.

Cattinelli, I., Borghese, N.A., Gallucci, M., Paulesu, E., 2013. Reading the reading brain. A new meta-analysis of functional imaging data on reading. J. Neurolinguistics 26 (1), 214–238. https://doi.org/10.1016/j.jneuroling.2012.08.001.

Chen, J.L., Rae, C., Watkins, K.E., 2012. Learning to play a melody. An fMRI study examining the formation of auditory-motor associations. NeuroImage 59 (2), 1200–1208. https://doi.org/10.1016/j.neuroimage.2011.08.012.

Choi, I., Lee, J.Y., Lee, S.H., 2018. Bottom-up and top-down modulation of multisensory integration. Curr. Opin. Neurol. 52, 115–122. https://doi.org/10.1016/j.conb.2018.05.002.

DeMarco, A.T., Wilson, S.M., Rising, K., Rapcsak, S.Z., Beeson, P.M., 2017. Neural substrates of sublexical processing for spelling. Brain Lang. 164, 118–128. https://doi.org/10.1016/j.bandl.2016.10.001.

Doehrmann, O., Naumer, M.J., 2008. Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res. 1242, 136–150. https://doi.org/10.1016/j.brainres.2008.03.071.

Ehrsson, H., Henrik, H., Nicholas, P., Passingham, R.E., 2005. Touching a rubber hand. Feeling of body ownership is associated with activity in multisensory brain areas. J. Neurosci. 25 (45), 10564–10573. https://doi.org/10.1523/JNEUROSCI.0800-05.2005.

Fatma, D., Nunez-Elizalde, A.O., Huth, A.G., Gallant, J.L., 2019. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. J. Neurosci. 39 (39), 7722–7736. https://doi.org/10.1523/JNEUROSCI.0675-19.2019.

Hein, G., Doehrmann, O., Müller, N.G., Kaiser, J., Muckli, L., Naumer, M.J., 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. J. Neurosci. 27 (30), 7881–7887. https://doi.org/10.1523/JNEUROSCI.1740-07.2007.

Junker, F.B., Schlaffke, L., Bellebaum, C., Ghio, M., Brühl, S., Axmacher, N., Schmidt-Wilcke, T., 2020. Transition From Sublexical to Lexico-Semantic Stimulus Processing. Front. Syst. Neurosci. 14, 129. https://doi.org/10.3389/fnsys.2020.522384.

Klemen, J., Chambers, C.D., 2012. Current perspectives and methods in studying neural mechanisms of multisensory interactions. Neurosci. Biobehav. Rev. 36 (1), 111–133. https://doi.org/10.1016/j.neubiorev.2011.04.015.

von Kriegstein, K., Giraud, A.-L., 2006. Implicit multisensory associations influence voice recognition. PLoS biology. 4 (10) https://doi.org/10.1371/journal.pbio.0040326.

Lakatos, P., Chen, C.-M., O'Connell, M.N., Mills, A., Schroeder, C.E., 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53 (2), 279–292. https://doi.org/10.1016/j.neuron.2006.12.011.

Lappe, C., Trainor, L.J., Herholz, S.C., Pantev, C., Zhang, L.I., 2011. Cortical plasticity induced by short-term multimodal musical rhythm training. PloS one 6 (6), e21493. https://doi.org/10.1371/journal.pone.0021493.

Lappe, C., Herholz, S.C., Trainor, L.J., Pantev, C., 2008. Cortical plasticity induced by short-term unimodal and multimodal musical training. J. Neurosci. 28 (39), 9632–9639. https://doi.org/10.1523/JNEUROSCI.2254-08.2008.

Lehmann, S., Murray, M.M., 2005. The role of multisensory memories in unisensory object discrimination. Brain Res. Cogn. Brain Res. 24 (2), 326–334. https://doi.org/10.1016/j.cogbrainres.2005.02.005.

Matusz, P.J., Thelen, A., Amrein, S., Geiser, E., Anken, J., Murray, M.M., 2015. The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. Eur. J. Neurosci. 41 (5), 699–708. https://doi.org/10.1111/ejn.12804.

Matusz, P.J., Wallace, M.T., Murray, M.M., 2017. A multisensory perspective on object memory. Neuropsychologia. 105, 243–252. https://doi.org/10.1016/j.neuropsychologia.2017.04.008.

Mcgurk, H., Macdonald, J., 1976. Hearing lips and seeing voices. Nature 264 (5588), 746–748.

McIntosh, A.R., Cabeza, R.E., Lobaugh, N.J., 1998. Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. J. Neurophysiol. 80 (5), 2790–2796. https://doi.org/10.1152/jn.1998.80.5.2790.

Naumer, M.J., Doehrmann, O., Müller, N.G., Muckli, L., Kaiser, J., Hein, G., 2009. Cortical plasticity of audio-visual object representations. Cereb. Cortex 19 (7), 1641–1653. https://doi.org/10.1093/cercor/bhn200.

Noppeney, U., Josephs, O., Hocking, J., Price, C.J., Friston, K.J., 2008. The effect of prior visual information on recognition of speech and sounds. Cereb. Cortex 18 (3), 598–609. https://doi.org/10.1093/cercor/bhm091.

Nyberg, L., Habib, R., McIntosh, A.R., Tulving, E., 2000. Reactivation of encoding-related brain activity during memory retrieval. Proc. Natl. Acad. Sci. U.S.A. 97 (20), 11110–11124. https://doi.org/10.1073/pnas.97.20.11120.

ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M.M., Matusz, P.J., 2016. The COGs (context, object, and goals) in multisensory processing. Exp. Brain Res. 234 (5), 1307–1323. https://doi.org/10.1007/s00221-016-4590-z.

Özyürek, A., 2014. Hearing and seeing meaning in speech and gesture. Insights from brain and behaviour. Philos. Trans. Royal Soc. B. 369 (1651), 20130296. https://doi.org/10.1098/rstb.2013.0296.

Paraskevopoulos, E., Herholz, S., 2013. Multisensory integration and neuroplasticity in the human cerebral cortex. Transl. Neurosci. 4 (3), 278. https://doi.org/10.2478/s13380-013-0134-1.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S.C., Pantev, C., Balasubramaniam, R., 2012. Evidence for training-induced plasticity in multisensory brain structures: an MEG study. PLoS One 7 (5), e36534. https://doi.org/10.1371/journal.pone.0036534.

Protopapas, A., Orfanidou, E., Taylor, J.S.H., Karavasilis, E., Kapnoula, E.C., Panagiotaropoulou, G., Velonakis, G., Poulou, L.S., Smyrnis, N., Kelekis, D., 2016. Evaluating cognitive models of visual word recognition using fMRI: effects of lexical and sublexical variables. NeuroImage 128, 328–341. https://doi.org/10.1016/j.neuroimage.2016.01.013.

Rowe, M.L., Goldin-Meadow, S., 2009. Differences in early gesture explain SES disparities in child vocabulary size at school entry. Science 323 (5916), 951–953. https://doi.org/10.1126/science.1167025.

Sadaghiani, S., Maier, J.X., Noppeney, U., 2009. Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. J. Neurosci. 29 (20), 6490–6499. https://doi.org/10.1523/JNEUROSCI.5437-08.2009.

Saldern, S., Noppeney, U., 2013. Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. J. Neurosci. 33 (20), 8841–8849. https://doi.org/10.1523/JNEUROSCI.3020-12.2013.

Schlaffke, L., Rüther, N.N., Heba, S., Haag, L.M., Schultz, T., Rosengarth, K., Tegenthoff, M., Bellebaum, C., Schmidt-Wilcke, T., 2015. From perceptual to lexico-semantic analysis–cortical plasticity enabling new levels of processing. Hum. Brain Mapp. 36 (11), 4512–4528. https://doi.org/10.1002/hbm.v36.1110.1002/hbm.22939.

Schlaffke, L., Leemans, A., Schweizer, L.M., Ocklenburg, S., Schmidt-Wilcke, T., 2017. Learning morse code alters microstructural properties in the inferior longitudinal fasciculus: a DTI study. Front. Hum. Neurosci. 11 (383) https://doi.org/10.3389/fnhum.2017.00383.

Seitz, A.R., Dinse, H.R., 2007. A common framework for perceptual learning. Curr. Opin. Neurobiol. 17 (2), 148–153. https://doi.org/10.1016/j.conb.2007.02.004.

Seitz, A.R., Kim, R., Shams, L., 2006. Sound facilitates visual learning. Curr. Biol. 16 (14), 1422–1427. https://doi.org/10.1016/j.cub.2006.05.048.

Scholz, J., Klein, M.C., Behrens, T.E.J., Johansen-Berg, H., 2009. Training induces changes in white-matter architecture. Nat. Neurosci. 12 (11), 1370–1371. https://doi.org/10.1038/nn.2412.

Shams, L., Seitz, A.R., 2008. Benefits of multisensory learning. Trends in Cogn. Sci. 12 (11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006.

Shams, L., Wozny, D.R., Kim, R., Seitz, A., 2011. Influences of multisensory experience on subsequent unisensory processing. Front. Psychol. 2 (264) https://doi.org/10.3389/fpsyg.2011.00264.

Thelen, A., Cappe, C., Murray, M.M., 2012. Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. NeuroImage 62 (3), 1478–1488. https://doi.org/10.1016/j.neuroimage.2012.05.027.

Taylor, J.S.H., Rastle, K., Davis, M.H., 2013. Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. Psychol. Bull. 139 (4), 766–791. https://doi.org/10.1037/a0030266.

Vigliocco, G., Perniss, P., Vinson, D., 2014. Language as a multimodal phenomenon. Implications for language learning, processing and evolution. Philos. Trans. Royal Soc. B. 369 (1651), 20130292. https://doi.org/10.1098/rstb.2013.0292.

Werner, S., Noppeney, U., 2010. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. 30 (7), 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010.

Xi, Y., Li, Q., Gao, N., He, S., Tang, X., Chen, K., 2019. Cortical network underlying audiovisual semantic integration and modulation of attention. An fMRI and graph-based study. PloS One 14 (8). https://doi.org/10.1371/journal.pone.0221185.

Yildirim, I., Jacobs, R.A., 2013. Transfer of object category knowledge across visual and haptic modalities. Experimental and computational studies. Cognition 126 (2), 135–148. https://doi.org/10.1016/j.cognition.2012.08.005.