

# PSYCHONOMIC BULLETIN & REVIEW

## Sensorimotor Training Modulates Automatic Imitation of Visual Speech

Journal:	<i>Psychonomic Bulletin &amp; Review</i>
Manuscript ID	PBR-BR-19-022.R3
Manuscript Type:	Brief Report
Date Submitted by the Author:	n/a
Complete List of Authors:	Wu, Yuchunzi; University College London, Department of Speech, Hearing & Phonetic Sciences Evans, Bronwen; University College London, Speech Hearing and Phonetic Sciences Adank, Patti; UCL, Speech, Hearing and Phonetic Sciences
Keywords:	automatic imitation, speech perception, speech production, sensorimotor learning

London, May 2019

Dear Dr. Catmur,

We would like to thank you for your comments and suggestions. In the revised manuscript, we changed the following sentence: “Our results do not allow us to completely exclude the possibility of an innate mechanism governing imitative, rather than counter- or non-imitative, sensorimotor associations underlying the imitation of orofacial movements (p.17).” Thanks for accepting our paper for publication in Psychonomic Bulletin & Review.

Yours sincerely,

Yuchunzi Wu (corresponding author), Bronwen G. Evans, Patti Adank.

FOR Review Only

Sensorimotor Training Modulates Automatic Imitation of Visual Speech

Yuchunzi Wu\*, Bronwen G. Evans, Patti Adank

Department of Speech, Hearing & Phonetic Sciences, University College London, London

WC1N 1PF, UK

\*Corresponding author at: Department of Speech, Hearing & Phonetic Sciences, University  
College London, London, WC1N 1PF, UK.

*Telephone number:* +44 (0)75 1179 8412

*Email address:* [yuchunzi.wu.15@ucl.ac.uk](mailto:yuchunzi.wu.15@ucl.ac.uk)

**Acknowledgements.** This work was supported by the China Scholarship Council to the first  
author and the BIAL Foundation under Grant No. 267/14 to P.A.

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

2

## Abstract

Observation-execution links underlying automatic imitation processes are suggested to result from associative sensorimotor experience of performing and watching the same actions. Past research supporting the associative sequence learning (ASL) model has demonstrated that sensorimotor training modulates automatic imitation of perceptually transparent manual actions, but ASL has been criticized for not being able to account for opaque actions like orofacial movements that include visual speech. To investigate whether observation-execution links underlying opaque actions are flexible as has been demonstrated for transparent actions, we tested whether sensorimotor training modulated automatic imitation of visual speech. Automatic imitation was defined as a facilitation in response times for syllable articulation (*ba* or *da*) in the presence of a compatible visual speech distractor relative to the presence of an incompatible distractor. Participants received either mirror (say /ba/ when the speaker silently says /ba/ and likewise for /da/) or counter-mirror (say /da/ when the speaker silently says /ba/ and vice versa) training and automatic imitation was measured before and after training. Automatic imitation was enhanced following mirror training and reduced following counter-mirror training, suggesting that sensorimotor learning plays a critical role in linking speech perception and production and that the links between these two systems remain flexible in adulthood. Additionally, compared to manual movements, automatic imitation of speech was susceptible to mirror training, but relatively resilient to counter-mirror training. We propose that social factors and the multimodal nature of speech may account for the resilience to counter-mirror training of sensorimotor associations of speech actions.

*Keywords:* automatic imitation, speech perception, speech production, sensorimotor learning

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

3

## Sensorimotor Training Modulates Automatic Imitation of Visual Speech

Observing an action activates the motor patterns used to perform the same action (Buccino et al., 2004; Hari et al., 1998), demonstrating an imitative capacity in the observer to quickly map observed actions onto his/her motor repertoire. The discovery of mirror neurons (MNs) in macaque monkeys (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) and humans (Mukamel, Ekstrom, Kaplan, Jacoboni, & Fried, 2010) demonstrated a direct neural observation-execution link, and neuroimaging studies have suggested a human mirror neuron system (MNS) that responds when participants execute and observe the same actions (Molenberghs, Cunnington, & Mattingley, 2012). The MNS has been proposed as the underlying neural structure sub-serving imitation (Buccino et al., 2004; Catmur, Walsh, & Heyes, 2009).

Behaviourally, observation-induced motor activation has been demonstrated in studies showing *automatic imitation*, measured using the stimulus-response compatibility (SRC) task (Heyes, Bird, Johnson, & Haggard, 2005; Stürmer, Aschersleben, & Prinz, 2000). In Stürmer et al., observing a compatible movement (e.g., an opening hand) facilitated participants' response (e.g., opening his/her hand) relative to observing an incompatible movement (e.g., a closing hand). Automatic imitation is defined as the response time (RT) difference between the two compatibility conditions, with a larger effect indicating greater observation-induced motor activation (Heyes, 2011). Automatic imitation is thought to occur because action observation activates corresponding motor patterns that interact with the participant's response. Specifically, performance is facilitated when observation activates the compatible action and is delayed when observation activates the incompatible action.

The associative sequence learning (ASL) model proposes that the imitative capacity is a product of associative sensorimotor learning that involves correlated experience of observing and

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

4

1  
2  
3 executing the same actions (Heyes, 2005, 2010). Notably, the associative mechanism is  
4  
5 suggested to be the same domain-general process that also produces Pavlovian and instrumental  
6  
7 conditioning and that is therefore sensitive to experienced stimulus-response pairs. Previous  
8  
9 studies supporting ASL demonstrated that sensorimotor training modulates automatic imitation  
10  
11 of manual movements (see Catmur, 2013 for review). In Heyes et al. (2005), automatic imitation  
12  
13 of hand opening/closing movements was eliminated following counter-mirror training that  
14  
15 associated different observed and executed hand movements, but not following mirror training  
16  
17 that associated the same observed and executed movements. Because both groups received the  
18  
19 same amount of sensory and motor practice during training, the authors concluded that it was the  
20  
21 relationship between observed and executed actions that modulated automatic imitation, hence  
22  
23 supporting ASL's hypothesis that observation-execution links depend on sensorimotor learning.  
24  
25  
26  
27

28  
29 Past training studies have exclusively examined perceptually transparent actions, such as  
30  
31 manual gestures, whose sensorimotor links could be built through self-observation, whereas one  
32  
33 dispute that remains unsolved concerns flexibility of links underlying perceptually opaque  
34  
35 orofacial gestures (Heyes, 2005). Based on the interpretation of infant imitation research, "innate  
36  
37 observation-execution links" (p. 23) are suggested to enable new-borns to imitate observed  
38  
39 orofacial actions (Meltzoff, 2002). Different developmental trajectories have also been  
40  
41 postulated for the manual and orofacial MNS. Specifically, Casile, Caggiano, & Ferrari (2011)  
42  
43 have suggested that the orofacial MNS is "prewired and already present at birth" (p. 532),  
44  
45 whereas the manual MNS is acquired after birth through learning. In contrast, ASL suggests that  
46  
47 observation-execution links underlying either manual or orofacial actions do not solely depend  
48  
49 on visual guidance of self-generated movements. Rather, imitative sensorimotor experience  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

5

1  
2  
3 mostly originates from sociocultural sources during development (e.g., being imitated by others  
4  
5 or through a common stimulus, Ray & Heyes, 2011).  
6

7  
8 This study aimed to determine the role of sensorimotor learning in establishing  
9  
10 observation-execution links underlying perceptually opaque orofacial gestures and we focused  
11  
12 on automatic imitation of visual speech. Speech actions are communicative orofacial gestures  
13  
14 seen in face-to-face conversations but not by talkers themselves. Watching and/or hearing other  
15  
16 people speak activates articulatory motor regions (Skipper, Devlin, & Lametti, 2017), suggesting  
17  
18 a close perception-production link. We aimed to elucidate whether the flexibility of observation-  
19  
20 execution links underlying manual gestures extends to speech perception-production links.  
21  
22

23  
24 Studies using speech SRC tasks have demonstrated that perceiving compatible  
25  
26 articulations produced by a speaker facilitates participants' responses relative to perceiving  
27  
28 incompatible articulations (Adank, Nuttall, Bekkering, & Maegherman, 2018; Kerzel &  
29  
30 Bekkering, 2000). Here, we adopted the speech SRC task to establish participants' initial  
31  
32 automatic imitation before assigning them to either a counter-mirror (say /ba/ when the speaker  
33  
34 says /da/ and vice versa) or a mirror (say /ba/ when the speaker says /ba/ and likewise for /da/)  
35  
36 training group. Automatic imitation was measured again using the same task 24 hrs after  
37  
38 training. We predicted that if sensorimotor experience is critical in establishing speech  
39  
40 perception-production links, automatic imitation would be impaired following counter-mirror  
41  
42 training but not following mirror training; however, if sensorimotor experience is not critical, we  
43  
44 predicted no difference in automatic imitation between two groups after training. Additionally, as  
45  
46 automatic imitation of speech has been shown to vary when prompts are presented at different  
47  
48 time points relative to distractor onsets (i.e., stimulus-onset asynchronies [SOAs]; Adank et al.,  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

6

2018), we included different SOAs in the SRC task to examine whether training effects would interact with automatic imitation at different SOAs.

### Method

#### Participants

An a-priori power analysis was conducted using G\*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) with an effect size  $\eta_p^2 = .108$ , obtained from a pilot study. Sixty-two participants were needed to detect significant sensorimotor training effects on automatic imitation with a power of .80 and an alpha of .05. Sixty-eight participants were recruited, but one was excluded for not being a native British English speaker, one for having dyslexia, one for not attending the post-training session and three for performing at chance level during training. The final analysis included 31 participants in the mirror group (23 female,  $M_{age} = 21.71$ ,  $SD_{age} = 4.89$ ,  $range_{age} = 17-34$ ) and 31 in the counter-mirror group (19 female,  $M_{age} = 21.61$ ,  $SD_{age} = 3.25$ ,  $range_{age} = 18-30$ ). All were native British English speakers with self-reported normal or corrected-to-normal vision, normal hearing and no speech or language disorders, or other neurological disorders. Participants received £20 or course credit. The University Research Ethics Committee approved the procedures and all participants gave written informed consent.



## SENSORIMOTOR MODULATION OF SPEECH IMITATION

7

## Stimuli and procedure

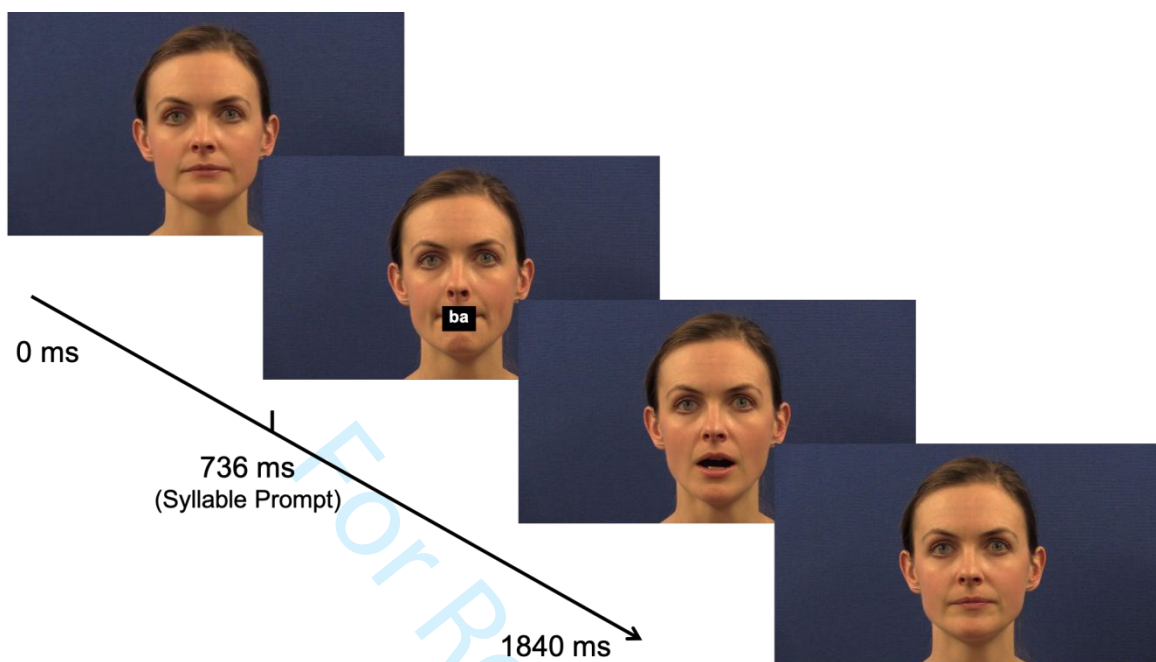


Fig. 1. Schematic timeline of a compatible trial presented in the testing sessions with the speaker saying /ba/ in the distractor video and the prompt *ba* appearing in the front. The video lasted 1840ms and the prompt was presented at 736ms. In the actual experiment, participants' viewing was unrestrained at a distance of 60 cm from the screen, and the speaker's face was shown in  $14.34^\circ \times 11.14^\circ$  of visual angle and the prompt was in  $0.38^\circ \times 0.38^\circ$  (see Supplementary Material for detailed measures of the speaker's mouth configurations at different time points). The size of the prompt is enlarged here for visual clarification.

Stimuli (Fig. 1) included silent videos of a speaker saying /ba/ and /da/ and syllable prompts *ba* and *da*. The videos (25 fps) were filmed with a Canon Legria HF G30 video camera, edited in iMovie and scaled down in resolution to 1280 x 720 in AVI format. A female native British English speaker was shown in the videos from her neckline upward in color. Both videos started and ended with the speaker's mouth closed in a resting configuration. At 552ms, the

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

8

1  
2  
3 speaker's mouth began to move in the *da* video, and the lips began moving towards each other in  
4  
5 the *ba* video. Consonant bursts in both videos occurred around 736ms and vowel articulation  
6  
7 commenced around 920ms. The speaker was still articulating the vowel at 1104ms and her facial  
8  
9 expression returned to its resting position at 1400ms. The prompts *ba* and *da* (300 dpi JPEG  
10  
11 images) were printed in white boldfaced Arial font on a black background and positioned  
12  
13 extending from the speaker's bottom lip to her top lip. The prompt was presented at one of four  
14  
15 SOAs (552, 736, 920 or 1104ms) in each trial. The experiment was performed using Presentation  
16  
17 (Version 18.0, Neurobehavioral Systems).  
18  
19

20  
21 The experiment included two testing sessions (pre- and post-training) and one training  
22  
23 session and took place in a soundproofed, light-controlled booth. Written instructions were  
24  
25 presented on the PC monitor. In the pre-training session, participants were instructed to speak out  
26  
27 the syllable (*ba* or *da*) as soon as they saw the prompt and to ignore the speaker's articulation (*ba*  
28  
29 or *da*) in the distractor video. In the compatible condition, the speaker's articulation matched the  
30  
31 prompted response; in the incompatible condition, the speaker's articulation differed from the  
32  
33 prompted response. Each trial started with a 200-ms tone with a frequency of 500 Hz at 70 dB  
34  
35 SPL played through Sennheiser HD25-SP II headphones. The screen remained black for one of  
36  
37 three jittered durations (1500, 1750 or 2000ms) that were included to reduce the trial onset's  
38  
39 predictability. The prompt was presented at one of the four SOAs for 200ms, and the screen went  
40  
41 black at the end of the video. There were six blocks with 40 trials each (240 trials in total) in the  
42  
43 pre-training session. Forty-eight trial types (2 Prompts x 2 Distractors x 4 SOAs x 3 Jitters) were  
44  
45 repeated five times in a randomized order. Ten practice trials were given before the first block  
46  
47 and the pre-training session lasted about 20min. The post-training session was identical to the  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

9

pre-training session, except that it was conducted the day after the pre-training session and that participants completed five practice trials.

Training took place immediately after the pre-training session and participants were randomly assigned to a training group. Participants in the counter-mirror group said /ba/ as soon as they saw the speaker mouth /da/ and vice versa; participants in the mirror group repeated the syllable as mouthed by the speaker. The same jittered inter-trial intervals were used. There were twelve blocks with 80 trials each (960 trials in total) and six trial types (2 Videos x 3 Jitters) were repeated in a randomized order. After the first six blocks, a short animation film was played with sound before they continued to the second half of the session. Five practice trials were given, and the training session lasted about 90 min.

**Data recording and analysis**

Responses were recorded via a voice-key using a RØDE NT1-A Condenser Microphone and a Focusrite Scarlett 2i4 USB Computer Audio Interface pre-amplifier plugged into the sound card input of a Dell PC at 44.1kHz with 16 bits. Audio recording started at the video onset for 3000ms. The voice-key was triggered when the system detected an audio input at .2 of Presentation's total range. RTs were measured relative to prompt onset. For missed trials, a warning saying *No response given* was presented for 500ms. A warning saying *Response too early* appeared for RTs <200ms. Responses were checked manually using Praat (Boersma & Weenink, 2018). Errors included incorrect responses and missed trials. For testing sessions, trials with RTs <100ms or >1200ms were defined as errors because they were likely to be anticipatory or neglected responses (Kerzel & Bekkering, 2000). For training sessions, outliers included trials with RTs that were three standard deviations away from the average. A natural log-

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

10

1  
2  
3 transformation was applied to RTs for statistical analyses, but figures present untransformed  
4  
5 RTs.  
6

7  
8 Error rates (ERs) and RTs from testing sessions were subjected to repeated-measures  
9  
10 Analysis of Variance (ANOVA) with test (pre- vs. post-training), compatibility (compatible vs.  
11  
12 incompatible) and SOA (552, 736, 920 or 1104ms) as within-subjects variables and training  
13  
14 (mirror vs. counter-mirror) as a between-subjects variable. ERs and RTs from training sessions  
15  
16 were analyzed in separate ANOVAs with block as a within-subjects variable and training as a  
17  
18 between-subjects variable. The significance level was set to  $p < .05$ . Greenhouse-Geisser  
19  
20 correction for non-sphericity and Bonferroni-correction for multiple comparisons were applied  
21  
22 whenever appropriate.  
23  
24

### 25 26 Results

27  
28 On average, participants made 7% errors in testing sessions. RT analyses for testing  
29  
30 sessions are reported here and other analyses are included in the Supplementary Materials. After  
31  
32 errors were excluded, RT analyses (Fig. 2, Table 1) revealed a main compatibility effect with  
33  
34 faster RTs for compatible trials ( $M = 6.226$ ,  $SE = .015$ ) than for incompatible trials ( $M = 6.303$ ,  
35  
36  $SE = .014$ ). There was a main test effect with slower RTs in the pre-training session ( $M = 6.302$ ,  
37  
38  $SE = .019$ ) than in the post-training session ( $M = 6.227$ ,  $SE = .013$ ). Follow-up t-tests for the  
39  
40 main effect of SOA revealed faster RTs for later SOAs (all  $p < .001$ ). For the interaction between  
41  
42 test and SOA, follow-up t-tests revealed greater RT reduction after training at three later SOAs  
43  
44 than at the first SOA (all  $p < .001$ ). Follow-up t-tests for the interaction between compatibility  
45  
46 and SOA revealed larger compatibility effects at two later SOAs than at two earlier SOAs (all  $p$   
47  
48  $< .003$ ).  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

11

1  
2  
3 A significant three-way interaction was found between training, test and compatibility;  
4 follow-up t-tests revealed that compatibility effects increased after mirror training ( $p = .002$ ) but  
5 did not change after counter-mirror training ( $p = .177$ ). This three-way interaction was further  
6 modulated by SOA, as suggested by the significant four-way interaction between training, test,  
7 compatibility and SOA. Follow-up t-tests of the four-way interaction (Fig. 2, Table 2) revealed  
8 that automatic imitation increased by 19ms from 42ms at each of two later SOAs after mirror  
9 training (920-ms:  $t(30) = 3.25$ ; 1104-ms:  $t(30) = 3.03$ ; all  $p < .006$ ) and decreased by 16ms from  
10 49ms at the last SOA after counter-mirror training (1104-ms:  $t(30) = 2.07$ ,  $p = .047$ ).  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

12

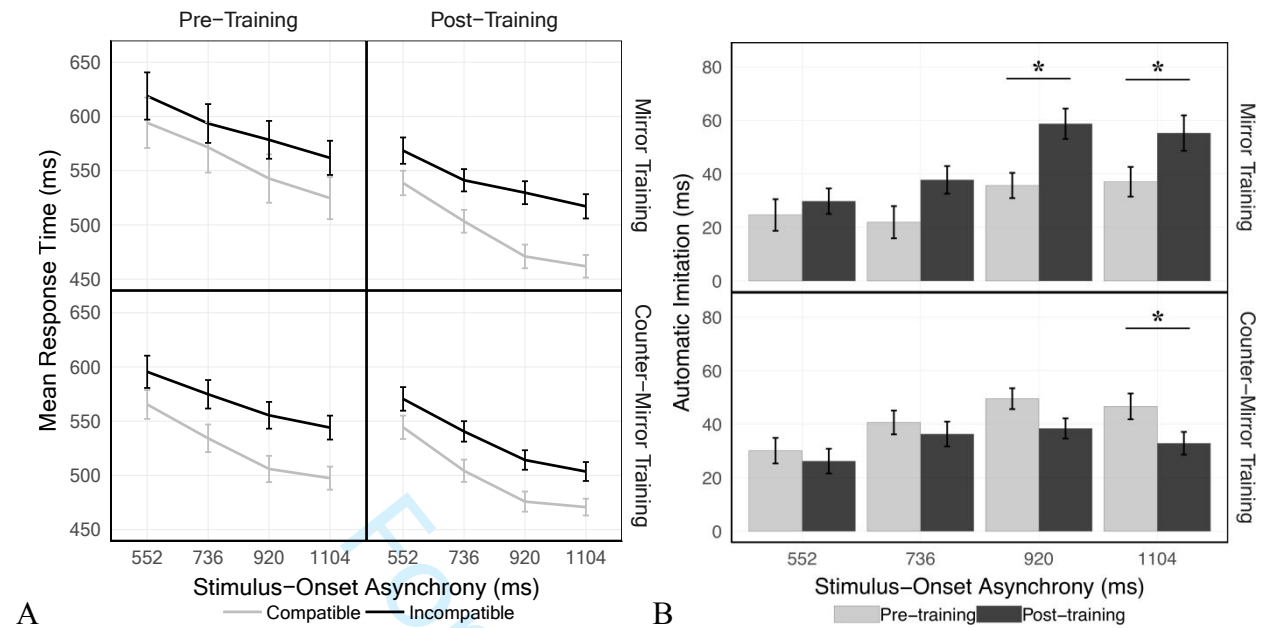


Fig. 2. (A) Mean response time  $\pm$  standard error in each experimental condition. Four panels represent pre- (top-left) and post-training (top-right) sessions in the mirror group and pre- (bottom-left) and post-training (bottom-right) sessions in the counter-mirror group. (B) Automatic imitation effects (i.e., incompatible - compatible)  $\pm$  standard error for pre- (grey) and post-training (black) sessions at each stimulus-onset asynchrony in each group. \* highlights significant changes in automatic imitation effects after training ( $p < .05$ ).

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

13

Table 1

*Four-Way ANOVA Summary for Response Times by Training, Test, Compatibility and Stimulus-Onset Asynchrony (SOA)*

Main Effect/Interaction	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Training	1, 60	.41	.524	.007
Test	1, 60	31.20	< .001	.342
Compatibility	1, 60	180.70	< .001	.751
SOA	1.68, 100.52	212.22	< .001	.780
Training x Test	1, 60	2.53	.117	.041
Training x Compatibility	1, 60	.26	.609	.004
Training x SOA	1.68, 100.52	.68	.486	.011
Test x Compatibility	1, 60	3.56	.064	.056
Test x SOA	2.59, 155.35	9.67	< .001	.139
Compatibility x SOA	2.31, 138.81	18.92	< .001	.240
Training x Test x Compatibility	1, 60	12.99	.001	.178
Training x Test x SOA	2.59, 155.35	.04	.984	.001
Training x Compatibility x SOA	2.31, 138.81	1.61	.199	.026
Test x Compatibility x SOA	3, 180	.85	.465	.014
Training x Test x Compatibility x SOA	3, 180	3.42	.019	.054

*Note:* Highlighted rows = significant effects/interactions with  $p < .05$ .

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

14

Table 2

Mean log-transformed response time, back-transformed response time in ms, standard error and 95% confidence interval for each experimental condition in the testing sessions.

Group		SOA	<i>M</i>	RT(ms)	<i>SE</i>	95% <i>CI</i>	
Mirror	Pre-training	Compatible	1	6.355	575	.030	[6.296, 6.414]
			2	6.314	552	.030	[6.254, 6.374]
			3	6.262	524	.031	[6.201, 6.323]
			4	6.232	509	.027	[6.179, 6.286]
		Incompatible	1	6.401	602	.029	[6.343, 6.458]
			2	6.363	580	.026	[6.312, 6.414]
			3	6.339	566	.025	[6.290, 6.388]
			4	6.311	551	.023	[6.265, 6.357]
	Post-training	Compatible	1	6.272	530	.020	[6.233, 6.312]
			2	6.200	493	.021	[6.159, 6.241]
			3	6.131	460	.022	[6.087, 6.175]
			4	6.104	448	.020	[6.064, 6.145]
		Incompatible	1	6.326	559	.020	[6.286, 6.366]
			2	6.280	534	.018	[6.243, 6.317]
			3	6.256	521	.019	[6.217, 6.295]
			4	6.231	508	.020	[6.191, 6.272]
Counter-Mirror	Pre-training	Compatible	1	6.318	554	.030	[6.258, 6.377]
			2	6.261	524	.030	[6.201, 6.321]
			3	6.205	495	.031	[6.144, 6.266]
			4	6.186	486	.027	[6.133, 6.240]
		Incompatible	1	6.368	583	.029	[6.310, 6.425]
			2	6.335	564	.026	[6.284, 6.386]
			3	6.301	545	.025	[6.251, 6.350]
			4	6.282	535	.023	[6.236, 6.328]
	Post-training	Compatible	1	6.286	537	.020	[6.246, 6.326]
			2	6.206	496	.021	[6.165, 6.248]
			3	6.145	466	.022	[6.102, 6.189]
			4	6.138	463	.020	[6.097, 6.178]
		Incompatible	1	6.332	562	.020	[6.292, 6.372]
			2	6.280	534	.018	[6.243, 6.316]
			3	6.230	508	.019	[6.191, 6.269]
			4	6.207	496	.020	[6.167, 6.247]

Note: SOA = stimulus-onset asynchrony; SOA1 = 552ms; SOA2 = 736ms; SOA3 = 920ms;

SOA4 = 1104ms; RT = back-transformed response time.



### Discussion

This study investigated sensorimotor training effects on automatic imitation of visual speech. We found that automatic imitation increased after mirror training and decreased after counter-mirror training. Moreover, mirror training had stronger effects than counter-mirror training. Our findings are largely consistent with the ASL hypothesis that observation-execution links underlying orofacial movements such as speech can be modulated through sensorimotor learning, therefore suggesting similar developmental trajectories for perceptually opaque and transparent actions.

ASL proposes that sensorimotor experience of observing and executing the same action establishes and strengthens excitatory matching links between sensory and motor representations of that action (Press, Gillmeister, & Heyes, 2007). Accordingly, mirror training in this study strengthened the excitatory matching links that consequently enhanced observation-induced motor activation, leading to more facilitation in the compatible condition relative to the incompatible condition (i.e., increased automatic imitation). ASL also proposes that sensorimotor experience of observing and executing different actions leads to excitatory non-matching links between sensory and motor representations of different actions and also establishes inhibitory matching links between sensory and motor representations of the same actions (Heyes et al., 2005). Accordingly, counter-mirror training in this study established inhibitory matching links that consequently reduced observation-induced motor activation, leading to less facilitation in the compatible condition relative to the incompatible condition (i.e., decreased automatic imitation).

Training effects were only found for later SOAs, where automatic imitation was larger. Speech actions consist of sequences of movements, and SOAs have been included in speech SRC tasks to demonstrate the time course of automatic imitation influenced by different

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

16

1  
2  
3 movement components of perceived speech. Potentially, larger automatic imitation at later SOAs  
4  
5 was elicited by perceptually more salient components of perceived actions and participants also  
6  
7 paid more attention to these components during training. Consequently, automatic imitation  
8  
9 elicited by these components was more susceptible to training. Future research could manipulate  
10  
11 participants' attention to different action components during training and examine whether such  
12  
13 manipulation influences training effects.  
14  
15

16  
17 Importantly, automatic imitation of speech seems more resilient to counter-mirror  
18  
19 training than automatic imitation of manual actions. In Heyes et al. (2005), automatic imitation  
20  
21 of manual movements was eliminated following counter-mirror training. However, automatic  
22  
23 imitation of speech actions in our study was only reduced after counter-mirror training (960 trials  
24  
25 in total) that was considerably longer than the training in Heyes et al. (432 trials). Following  
26  
27 ASL, this result may be accounted for by sociocultural imitative experience; it is possible that  
28  
29 sensorimotor experience of observing and executing the same orofacial movements is mostly  
30  
31 gained through social interactions. In monkeys, mouth MNs are found to be connected to brain  
32  
33 regions involved in emotion/reward processing that plays a role in social activities (Ferrari,  
34  
35 Gerbella, Coudé, & Rozzi, 2017). Comparable mirror activation has also been found in  
36  
37 homologous regions in humans during perception and production of emotional facial expressions  
38  
39 (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003). Hence, while both manual and orofacial  
40  
41 observation-execution links are likely to result from sensorimotor learning, the extent of social  
42  
43 influence may differ between the two with the latter requiring more social engagement. Future  
44  
45 studies could investigate how social manipulations modulate sensorimotor training effects on  
46  
47 automatic imitation of manual and orofacial movements.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

17

1  
2  
3 Mirror training was more effective than counter-mirror training in our study, showing the  
4  
5 opposite of what has previously been reported (Cracco et al., 2018). ASL proposes that  
6  
7 observation-execution links could also be built through a common stimulus (e.g., hearing people  
8  
9 say /ba/) that co-occurs with experience of performing a movement (e.g., saying /ba/) and with  
10  
11 experience of seeing others perform the same movement (e.g., seeing people say /ba/) (Heyes,  
12  
13 2005). Such sensorimotor experience may contribute to the creation of indirect observation-  
14  
15 execution links different from the direct ones learned through the training provided in our study.  
16  
17 There may exist direct and indirect links acquired through different experience and the latter  
18  
19 could have been initially stronger for speech actions that are inherently multimodal. Therefore,  
20  
21 our finding was likely due to initially weak direct links that were more susceptible to mirror  
22  
23 training than to counter-mirror training in modulating automatic imitation of speech.  
24  
25  
26  
27

28 Our results do not allow us to completely exclude the possibility of an innate mechanism  
29  
30 governing imitative, rather than counter- or non-imitative, sensorimotor associations underlying  
31  
32 the imitation of orofacial movements. Heyes (2011) also acknowledges that results from training  
33  
34 studies in principle “do not exclude a role for genetic prespecification in establishing the long-  
35  
36 term sensorimotor connections that generate automatic imitation” (p. 478). Nevertheless, though  
37  
38 the current study does not conclusively support ASL, our results are in line with the main ASL  
39  
40 hypothesis that it is sensorimotor experience, but not sensory or motor experience alone, that  
41  
42 configures observation-execution links (Heyes, 2010), since the only difference between the two  
43  
44 groups in our study was the relationship between observed and executed movements per trial  
45  
46 during training. Hence, our findings suggest that the ASL mechanism can also be applied to  
47  
48 communicative orofacial movements that infants learn to perceive and produce in the first few  
49  
50 years of life. The precise mechanisms responsible for forging sensorimotor associations  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

18

1  
2  
3 underlying speech actions may be further explored by providing extended counter-mirror  
4 training. If extended training leads to a reduction/reversal of automatic imitation, this would  
5 support the notion that learning of these associations is not necessarily constrained by innate  
6 factors.  
7  
8  
9  
10

11  
12 The simulation theory of speech perception proposes that observation-induced motor  
13 activation facilitates prediction of the incoming signals supporting speech comprehension  
14 (Pickering & Garrod, 2013). Critically, greater motor involvement is suggested when observers  
15 have more experience with the perceived speech. Applying transcranial magnetic stimulation  
16 (TMS) to lip motor cortex (lip M1), Swaminathan et al. (2013) found facilitated lip M1  
17 excitability during the viewing of sentences spoken in a known language than in an unknown  
18 language. Following ASL, Swaminathan et al. suggested that the difference between two  
19 conditions was due to different strength of perception-production links underlying known and  
20 unknown languages, hence supporting the simulation hypothesis that more experience leads to  
21 greater observation-induced motor activation. Our results suggest that it was imitative  
22 sensorimotor learning that facilitated observation-induced motor activation. Additionally, overt  
23 imitation of accented speech improves subsequent speech perception, indicating that imitative  
24 learning leads to enhanced observation-induced motor activation facilitating speech  
25 comprehension (Adank, Hagoort, & Bekkering, 2010). Schmitz et al. (2018) stimulated lip M1  
26 with TMS and found that listening to non-native vowels elicited higher articulatory excitability  
27 than native-like vowels, which is opposite to what was found in Swaminathan et al. where  
28 sentence articulations were presented visually. Future research that controls linguistic levels and  
29 stimulus modalities is required to investigate this inconsistency. Moreover, follow-up research  
30 could extend our findings by examining whether sensorimotor training modulates audio-motor  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

19

1  
2  
3 links underlying speech. Behavioral research could also examine automatic imitation of non-  
4  
5 native speech and investigate the role of sensorimotor learning in establishing perception-  
6  
7 production links underlying second language processing.  
8  
9

10 In conclusion, the current study showed that sensorimotor training modulated automatic  
11  
12 imitation of visual speech. As such, our results elucidate questions concerning the flexibility of  
13  
14 the imitative mechanisms and adds to the growing body of evidence on perception-production  
15  
16 links in speech processing.  
17  
18

19  
20  
21 Open Practices Statements:

22 The data and materials for this experiment are available upon request and no experiment was  
23  
24 preregistered.  
25  
26  
27  
28  
29

30  
31 References

32  
33 Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension.

34  
35 *Psychological Science*, 21(12), 1903–1909.

36  
37 Adank, P., Nuttall, H., Bekkering, H., & Maegherman, G. (2018). Effects of stimulus response

38  
39 compatibility on covert imitation of vowels. *Attention, Perception, & Psychophysics*, 1–10.

40  
41 Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer. Retrieved from

42  
43 <http://www.praat.org/>

44  
45  
46 Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., & Rizzolatti, G. (2004).

47  
48 Neural circuits underlying imitation learning of hand actions: An event-related fMRI study.

49  
50 *Neuron*, 42(2), 323–334.

51  
52  
53 Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural

54  
55  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

20

1  
2  
3 mechanisms of empathy in humans: A relay from neural systems for imitation to limbic  
4 areas. *Proceedings of the National Academy of Sciences*, 100(9), 5497–5502.

5  
6  
7 Casile, A., Caggiano, V., & Ferrari, P. F. (2011). The mirror neuron system: A fresh view.  
8  
9  
10 *Neuroscientist*, 17(5), 524–538.

11  
12 Catmur, C. (2013). Sensorimotor learning and the ontogeny of the mirror neuron system.  
13  
14  
15 *Neuroscience Letters*, 540, 21–27.

16  
17 Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
experience in the development of imitation and the mirror system. *Philosophical  
Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1528),  
2369–2380.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., ... Brass, M. (2018).  
Automatic imitation: A meta-analysis. *Psychological Bulletin*, 144(5), 453–500.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using  
G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*,  
41(4), 1149–1160.

Ferrari, P. F., Gerbella, M., Coudé, G., & Rozzi, S. (2017). Two different mirror neuron  
networks: The sensorimotor (hand) and limbic (face) pathways. *Neuroscience*, 358, 300–  
315.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor  
cortex. *Brain*, 119(2), 593–609.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998).  
Activation of human primary motor cortex during action observation: A neuromagnetic  
study. *Proceedings of the National Academy of Sciences*, 95(25), 15061–15065.

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

21

- 1  
2  
3 Heyes, C. (2005). Imitation by Association. In S. Hurley & N. Chater (Eds.), *Perspectives on*  
4  
5 *Imitation: From Neuroscience to Social Science: Volume 1 Mechanisms of Imitation and*  
6  
7 *Imitation in Animals* (pp. 157–176). MIT Press.
- 8  
9  
10 Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral*  
11  
12 *Reviews*, 34(4), 575–583.
- 13  
14 Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, 137(3), 463–483.
- 15  
16  
17 Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic  
18  
19 imitation. *Cognitive Brain Research*, 22(2), 233–240.
- 20  
21  
22 Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from  
23  
24 stimulus response compatibility. *Journal of Experimental Psychology: Human Perception*  
25  
26 *and Performance*, 26(2), 634–647.
- 27  
28  
29 Meltzoff, A. N. (2002). Elements of a developmental theory of imitation. In A. N. Meltzoff & W.  
30  
31 Prinz (Eds.), *The Imitative Mind Development, Evolution and Brain Bases* (pp. 19–41).  
32  
33 Cambridge: Cambridge University Press.
- 34  
35  
36 Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror  
37  
38 properties: A meta-analysis of 125 human fMRI studies. *Neuroscience and Biobehavioral*  
39  
40 *Reviews*, 36(1), 341–349.
- 41  
42  
43 Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron  
44  
45 Responses in Humans during Execution and Observation of Actions. *Current Biology*,  
46  
47 20(8), 750–756.
- 48  
49  
50 Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and  
51  
52 comprehension. *Behavioral and Brain Sciences*, 36(04), 329–347.
- 53  
54  
55 Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic  
56  
57  
58  
59  
60

## SENSORIMOTOR MODULATION OF SPEECH IMITATION

22

1  
2  
3 imitation of robotic action. *Proceedings of the Royal Society B: Biological Sciences*,  
4  
5 274(1625), 2509–2514.  
6

7  
8 Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental*  
9  
10 *Science*, 14(1), 92–105.  
11

12 Schmitz, J., Bartoli, E., Maffongelli, L., Fadiga, L., Sebastian-Galles, N., & D’Ausilio, A.  
13  
14 (2018). Motor cortex compensates for lack of sensory and motor experience during auditory  
15  
16 speech perception. *Neuropsychologia*, (July 2017), 1–7.  
17  
18

19 Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the  
20  
21 speaking tongue : Review of the role of the motor system in speech perception. *Brain and*  
22  
23 *Language*, 164, 77–105.  
24  
25

26 Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures  
27  
28 and postures: a study of imitation. *Journal of Experimental Psychology: Human Perception*  
29  
30 *and Performance*, 26(6), 1746–1759.  
31  
32

33 Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., & Möttönen, R.  
34  
35 (2013). Motor excitability during visual perception of known and unknown spoken  
36  
37 languages. *Brain and Language*, 126(1), 1–7.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## “Sensorimotor training modulates automatic imitation of visual speech”

Yuchunzi Wu, Bronwen G. Evans, Patti Adank

### Supplemental results

#### Stimuli

In both videos, the speaker’s mouth was shown in  $0.99^\circ \times 3.81^\circ$  of visual angle at its resting configuration and in  $2.03^\circ \times 3.23^\circ$  during the vowel articulation. The most salient visual difference between two videos was before consonant bursts, with the speaker’s mouth shown in  $0.46^\circ \times 3.81^\circ$  in the *ba* video and in  $1.51^\circ \times 3.59^\circ$  in the *da* video.

#### Error rate analysis for testing sessions.

On average, participants made 7% errors (incorrect responses: 1.6%; missed trials: 1.6%; RTs < 100 ms: 1.7%; RTs > 1200 ms: 2.1%) in testing sessions. ER analyses (see Table S.1 and Table S.2) revealed a main compatibility effect, with a lower ER for compatible trials ( $M = 5.9$ ,  $SE = .008$ ) than for incompatible trials ( $M = 8$ ,  $SE = .010$ ). Planned t-tests for the interaction between test and SOA, revealed a slightly increased ER at the last SOA after training ( $p = .058$ ) relative to the ERs at the other three SOAs (all  $p > .3$ ). The interaction between training, test and SOA was significant, and follow-up t-tests revealed an increased ER at the last SOA after mirror training ( $p = .037$ ) but not after counter-mirror training ( $p = .831$ ).

Table S.1

*Four-Way ANOVA Summary for Error Rates by Training, Test, Compatibility, and Stimulus-*

*Onset Asynchrony (SOA)*

Main Effect/Interaction	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Training	1, 60	.65	.423	.011
Test	1, 60	.01	.937	<.001
Compatibility	1, 60	12.20	.001	.169
SOA	1.55, 92.73	1.58	.213	.026
Training x Test	1, 60	.25	.622	.004
Training x Compatibility	1, 60	1.34	.251	.017
Training x SOA	1.55, 92.73	.41	.614	.007
Test x Compatibility	1, 60	1.01	.318	.017
Test x SOA	1.76, 105.58	6.72	.003	.101
Compatibility x SOA	2.30, 138.13	.53	.616	.009
Training x Test x Compatibility	1, 60	.61	.439	.010
Training x Test x SOA	1.76, 105.58	4.04	.025	.063
Training x Compatibility x SOA	2.30, 138.12	1.31	.273	.021
Test x Compatibility x SOA	3, 18	.77	.512	.011
Training x Test x Compatibility x SOA	3, 18	.19	.900	.003

*Note:* Highlighted rows = significant effects/interactions with  $p < .05$ .

Table S.2

Mean error rate, standard error and 95% confidence interval for each experimental condition in the testing sessions.

Group	SOA	<i>M</i>	<i>SE</i>	95% <i>CI</i>		
Mirror	Pre-training	Compatible	1	8.1	2.5	[3.1, 13.1]
			2	6.5	2.1	[2.3, 10.6]
			3	5.6	1.7	[2.2, 9.0]
			4	4.8	1.2	[2.4, 7.3]
	Incompatible	1	9.4	2.2	[5.0, 13.7]	
		2	9.9	2.2	[5.5, 14.3]	
		3	7.3	1.9	[3.5, 11.1]	
		4	6.8	2.0	[2.7, 10.9]	
	Post-training	Compatible	1	4.8	1.2	[2.5, 7.2]
			2	5.5	1.1	[3.4, 7.6]
			3	6.2	1.1	[4.0, 8.4]
			4	8.7	1.6	[5.6, 11.9]
		Incompatible	1	7.6	1.5	[4.6, 10.6]
			2	8.8	1.4	[5.9, 11.7]
			3	10.3	1.8	[6.7, 14.0]
			4	11.9	2.3	[7.3, 16.5]
Counter-Mirror	Pre-training	Compatible	1	5.7	2.5	[0.7, 10.7]
			2	6.5	2.1	[2.3, 10.6]
			3	5.5	1.7	[2.1, 8.9]
			4	5.8	1.2	[3.4, 8.2]
	Incompatible	1	6.6	2.2	[2.2, 11.0]	
		2	6.4	2.2	[2.0, 10.8]	
		3	6.7	1.9	[2.9, 10.5]	
		4	8.9	2.0	[4.8, 13.0]	
	Post-training	Compatible	1	4.9	1.2	[2.6, 7.3]
			2	5.3	1.1	[3.2, 7.4]
			3	4.3	1.1	[2.1, 6.5]
			4	6.6	1.6	[3.4, 9.7]
		Incompatible	1	6.7	1.5	[3.7, 9.7]
			2	4.9	1.4	[2.1, 7.8]
			3	6.6	1.8	[2.9, 10.2]
			4	8.7	2.3	[4.1, 13.3]

Note: SOA = stimulus-onset asynchrony; SOA1 = 552 ms; SOA2 = 736 ms; SOA3 = 920 ms; SOA4 = 1104 ms.

### **Error rate and response time analysis for training sessions.**

Due to technical problems, one participant in the mirror group finished 650 trials and one in the counter-mirror group finished 921 trials. Errors (4.8%) included incorrect responses (3.1%) and missed trials (1.7%). ER analyses (see Table S.3) revealed a main training effect,  $F(1, 59) = 12.18, p = .001, \eta_p^2 = .171$ , with a higher ER for the counter-mirror group ( $M = 7.3, SE = .010$ ) than for the mirror group ( $M = 2.1, SE = .010$ ). No other effects were found. Errors and outliers (2.6%) were then excluded. RT analyses (see Table S.3) revealed a main training effect,  $F(1, 59) = 17.302, p < .001, \eta_p^2 = .227$ , with slower RTs in the counter-mirror group ( $M = 7.121, SE = .027$ ) than in the mirror group ( $M = 6.962, SE = .027$ ). There was a main effect of block,  $F(3.81, 224.51) = 17.77, p < .001, \eta_p^2 = .231$ , and follow-up tests revealed that RTs in the first three blocks were slower than those in the last four blocks (all  $p < .001$ ). The interaction between training and block was not significant.

Table S.3

*Mean error rate, mean log-transformed response time, back-transformed response time in ms and their corresponding standard errors and 95% confidence intervals for each block in the training sessions.*

Group	Block	Error Rate			Response Time			
		<i>M</i>	<i>SE</i>	95% <i>CI</i>	<i>M</i>	RT(ms)	<i>SE</i>	95% IC
Mirror	1	2.3	1.2	[-0.1, 4.7]	7.059	1163	.024	[7.011, 7.107]
	2	2.3	1.3	[-0.3, 4.9]	7.006	1103	.029	[6.948, 7.065]
	3	2.0	1.1	[-0.3, 4.2]	6.990	1086	.030	[6.929, 7.051]
	4	2.5	0.9	[0.7, 4.3]	6.969	1063	.030	[6.908, 7.030]
	5	1.9	1.3	[-0.6, 4.4]	6.951	1044	.032	[6.887, 7.015]
	6	2.7	1.2	[0.2, 5.1]	6.956	1049	.029	[6.899, 7.014]
	7	2.0	1.1	[-0.1, 4.2]	6.944	1037	.028	[6.887, 7.000]
	8	2.0	0.9	[0.1, 3.9]	6.947	1040	.028	[6.890, 7.004]
	9	2.2	1.2	[-0.2, 4.6]	6.932	1025	.028	[6.876, 6.989]
	10	2.2	1.3	[-0.4, 4.8]	6.922	1014	.032	[6.857, 6.986]
	11	1.8	1.2	[-0.7, 4.2]	6.938	1031	.031	[6.876, 7.000]
	12	1.7	1.4	[-1.1, 4.5]	6.926	1018	.036	[6.855, 6.997]
Counter-Mirror	1	6.3	1.2	[3.9, 8.6]	7.187	1322	.024	[7.140, 7.235]
	2	7.4	1.3	[4.8, 10.0]	7.171	1301	.029	[7.113, 7.229]
	3	7.2	1.1	[5.0, 9.4]	7.162	1289	.030	[7.102, 7.222]
	4	6.7	0.9	[5.0, 8.5]	7.146	1269	.030	[7.086, 7.206]
	5	8.1	1.2	[5.6, 10.6]	7.125	1243	.031	[7.062, 7.188]
	6	7.9	1.2	[5.4, 10.3]	7.121	1238	.028	[7.064, 7.178]
	7	6.2	1.1	[4.1, 8.3]	7.104	1217	.028	[7.048, 7.159]
	8	6.4	0.9	[4.5, 8.2]	7.116	1232	.028	[7.060, 7.172]
	9	7.2	1.2	[4.9, 9.6]	7.097	1208	.028	[7.041, 7.152]
	10	7.4	1.3	[4.8, 10.0]	7.073	1180	.032	[7.009, 7.136]
	11	8.5	1.2	[6.1, 10.9]	7.091	1201	.030	[7.030, 7.152]
	12	7.8	1.4	[5.0, 10.5]	7.056	1160	.035	[6.986, 7.126]

*Note:* RT = back-transformed response time.