Sensorimotor training modulates automatic imitation of visual speech

Yuchunzi Wu¹ · Bronwen G. Evans¹ · Patti Adank¹

Published online: 13 June 2019
© The Psychonomic Society, Inc. 2019

Abstract

The 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, such as orofacial movements that include visual speech. To investigate whether the observation-execution links underlying opaque actions are as flexible as has been demonstrated for transparent actions, we tested whether sensorimotor training modulated the automatic imitation of visual speech. Automatic imitation was defined as a facilitation in response times for syllable articulation (*ba* or *da*) when in the presence of a compatible visual speech distractor, relative to when in the presence of an incompatible distractor. Participants received either mirror (say /*ba*/ when the speaker silently says /*ba*/, and likewise for /*da*/) or countermirror (say /*da*/ when the speaker silently says /*ba*/, and vice versa) training, and automatic imitation was measured before and after training. The automatic-imitation effect was enhanced following mirror training and reduced following countermirror 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, as compared to manual movements, automatic imitation of speech was susceptible to mirror training but was relatively resilient to countermirror training. We propose that social factors and the multimodal nature of speech might account for the resilience to countermirror training of sensorimotor associations of speech actions.

Keywords Automatic imitation · Speech perception · Speech production · Sensorimotor learning

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 or her motor repertoire. The discovery of mirror neurons in macaque monkeys (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) and humans (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010) demonstrated a direct neural observation-execution link, and neuroimaging studies have suggested that a human mirror neuron system (MNS) responds when participants execute and observe the same actions (Molenberghs, Cunnington, & Mattingley,

2012). The MNS has been proposed as the underlying neural structure subserving imitation (Buccino et al., 2004; Catmur, Walsh, & Heyes, 2009).

Behaviorally, 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 their 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 the 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.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13423-019-01623-8>) contains supplementary material, which is available to authorized users.

✉ Yuchunzi Wu
yuchunzi.wu.15@ucl.ac.uk

¹ Department of Speech, Hearing & Phonetic Sciences, University College London, London, UK

The associative sequence learning (ASL) model proposes that the imitative capacity is a product of associative sensorimotor learning that involves correlated experiences of observing and executing the same actions (Heyes, 2005, 2010). Notably, the associative mechanism is suggested to be the same domain-general process that also produces Pavlovian and instrumental conditioning, and that is therefore sensitive to experienced stimulus–response pairs. Previous studies supporting ASL have demonstrated that sensorimotor training modulates automatic imitation of manual movements (see Catmur, 2013, for review). In Heyes et al. (2005), automatic imitation of hand opening/closing movements was eliminated following countermirror training that associated different observed and executed hand movements, but not following mirror training that associated the same observed and executed movements. Because both groups received the same amount of sensory and motor practice during training, the authors concluded that the relationship between observed and executed actions was what modulated automatic imitation, hence supporting ASL's hypothesis that observation–execution links depend on sensorimotor learning.

Past training studies have exclusively examined perceptually transparent actions, such as manual gestures, whose sensorimotor links could be built through self-observation, whereas one dispute that remains unresolved concerns the flexibility of the links underlying perceptually opaque orofacial gestures (Heyes, 2005). On the basis of the interpretation of infant imitation research, “innate observation–execution links” are suggested to enable newborns to imitate observed orofacial actions (Meltzoff, 2002, p. 23). Different developmental trajectories have also been postulated for the manual and orofacial MNS. Specifically, Casile, Caggiano, and Ferrari (2011) have suggested that the orofacial MNS is “prewired and already present at birth” (p. 532), whereas the manual MNS is acquired after birth through learning. In contrast, ASL suggests that the observation–execution links underlying either manual or orofacial actions do not depend solely on the visual guidance of self-generated movements. Rather, imitative sensorimotor experience mostly originates from sociocultural sources during development (e.g., being imitated by others, or through a common stimulus; Ray & Heyes, 2011).

In this study, we aimed to determine the role of sensorimotor learning in establishing the observation–execution links underlying perceptually opaque orofacial gestures, focusing on the automatic imitation of visual speech. Speech actions are communicative orofacial gestures seen in face-to-face conversations, but not by the talkers themselves. Watching and/or hearing other people speak activates articulatory motor regions (Skipper, Devlin, & Lametti, 2017), suggesting a close perception–production link. We aimed to elucidate whether the flexibility of the observation–execution links underlying

manual gestures extends to speech perception–production links.

Studies using speech SRC tasks have demonstrated that perceiving compatible articulations produced by a speaker facilitates participants' responses, relative to perceiving incompatible articulations (Adank, Nuttall, Bekkering, & Maegherman, 2018; Kerzel & Bekkering, 2000). Here we adopted the speech SRC task to establish participants' initial automatic imitation, before assigning them to either a countermirror (say /ba/ when the speaker says /da/, and vice versa) or a mirror (say /ba/ when the speaker says /ba/, and likewise for /da/) training group. Automatic imitation was measured again using the same task 24 h after training. We predicted that if sensorimotor experience is critical in establishing speech perception–production links, automatic imitation would be impaired following countermirror training, but not following mirror training; however, if sensorimotor experience is not critical, we predicted no difference in automatic imitation between the two groups after training. Additionally, because automatic imitation of speech has been shown to vary when prompts are presented at different time points relative to distractor onsets (i.e., stimulus onset asynchronies [SOAs]; Adank et al., 2018), we included different SOAs in the SRC task in order 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 of $\eta_p^2 = .108$, obtained from a pilot study. Sixty-two participants were needed in order 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 posttraining session, and three for performing at chance level during training. The final analysis included 31 participants in the mirror group (23 female, eight male; $M_{age} = 21.71$ years, $SD_{age} = 4.89$, range_{age} = 17–34) and 31 in the countermirror group (19 female, 12 male; $M_{age} = 21.61$ years, $SD_{age} = 3.25$, 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.

Stimuli and procedure

The stimuli (Fig. 1) included silent videos of a speaker saying /ba/ and /da/ and the 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 $1,280 \times 720$ in AVI format. A female native British English speaker was shown in color videos from her neckline upward. Both videos started and ended with the speaker's mouth closed in a resting configuration. At 552 ms, the speaker's mouth began to move in the *da* video, and the lips began moving toward each other in the *ba* video. The consonant bursts in both videos occurred around 736 ms, and the vowel articulation commenced around 920 ms. The speaker was still articulating the vowel at 1,104 ms, and her facial expression returned to its resting position at 1,400 ms. The prompts *ba* and *da* (300-dpi JPEG images) were printed in white boldface Arial font on a black background and were positioned extending from the speaker's bottom to top lip. The prompt was presented at one of four SOAs (552, 736, 920, or 1,104 ms) in each trial. The experiment was performed using Presentation (version 18.0, Neurobehavioral Systems).

The experiment included two testing sessions (pre- and posttraining) and one training session and took place in a soundproofed, light-controlled booth. Written instructions were presented on the PC monitor. In the pretraining session, participants were instructed to speak out the syllable (*ba* or *da*) as soon as they saw the prompt and to ignore the speaker's articulation (*ba* or *da*) in the distractor video. In the compatible

condition, the speaker's articulation matched the prompted response; in the incompatible condition, the speaker's articulation differed from the prompted response. Each trial started with a 200-ms tone with a frequency of 500 Hz at 70 dB SPL, played through Sennheiser HD25-SP II headphones. The screen remained black for one of three jittered durations (1,500, 1,750, or 2,000 ms), which were included to reduce the trial onset's predictability. The prompt was presented at one of the four SOAs for 200 ms, and the screen went black at the end of the video. There were six blocks with 40 trials each (240 trials in total) in the pretraining session. Forty-eight trial types ($2 \text{ prompts} \times 2 \text{ distractors} \times 4 \text{ SOAs} \times 3 \text{ jitters}$) were repeated five times in a randomized order. Ten practice trials were given before the first block, in a pretraining session that lasted about 20 min. The posttraining session was identical to the pretraining session, except that it was conducted the day after the pretraining session and participants completed five practice trials.

Training took place immediately after the pretraining session, and participants were randomly assigned to a training group. The participants in the countermirror group said /ba/ as soon as they saw the speaker mouth /da/, and vice versa when they said /da/; the participants in the mirror group repeated the syllable as it was mouthed by the speaker. The same jittered intertrial intervals were used. There were 12 blocks with 80 trials each (960 trials in total), in which the six trial types ($2 \text{ videos} \times 3 \text{ jitters}$) were presented in a randomized order. After the first six blocks, a short animation film was played with sound before participants continued to the second

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 1,840 ms, with the prompt presented at 736 ms. In the actual experiment, participants' viewing was unrestrained at a distance of 60 cm from the screen, the

speaker's face was shown at $14.34^\circ \times 11.14^\circ$ of visual angle, and the prompt measured $0.38^\circ \times 0.38^\circ$ (see the [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

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 preamplifier plugged into the sound card input of a Dell PC at 44.1 kHz with 16 bits. Audio recording started at the video onset for 3,000 ms. The voice key was triggered when the system detected an audio input at .2 of Presentation's total range. RTs were measured relative to the prompt onset. For missed trials, a warning stating "No response given" was presented for 500 ms. A warning stating "Response too early" appeared for RTs < 200 ms. The responses were checked manually using Praat (Boersma & Weenink, 2018). Errors included incorrect responses and missed trials. For the testing sessions, trials with RTs < 100 or > 1,200 ms were defined as errors, because they were likely to be anticipatory or neglected responses (Kerzel & Bekkering, 2000). For the training sessions, outliers included trials with RTs that were three standard deviations away from the average. A natural log-transformation was applied to the RTs for statistical analyses, but the figures present untransformed RTs.

Error rates and RTs from the testing sessions were subjected to repeated measures analyses of variance (ANOVAs) with test (pre- vs. posttraining), compatibility (compatible vs. incompatible), and SOA (552, 736, 920 or 1,104 ms) as within-subjects variables, and training (mirror vs. countermirror) as a between-subjects variable. Error rates and RTs from the training sessions were analyzed in separate ANOVAs, with block as a within-subjects variable and training as a between-subjects variable. The significance level was set to $p < .05$. Greenhouse–Geisser corrections for nonsphericity and Bonferroni corrections for multiple comparisons were applied whenever appropriate.

Results

On average, participants made 7% errors in the testing sessions. RT analyses for the testing sessions are reported here, and other analyses are included in the [supplementary materials](#). After errors were excluded, RT analyses (Fig. 2, Table 1) revealed a compatibility main effect, with faster log-transformed RTs for compatible trials ($M = 6.226$, $SE = 0.015$) than for incompatible trials ($M = 6.303$, $SE = 0.014$). We also found a test main effect, with slower log-transformed RTs in the pretraining session ($M = 6.302$, $SE = 0.019$) than in the posttraining session ($M = 6.227$, $SE = 0.013$). Follow-up t tests for the main effect of SOA revealed faster RTs for longer SOAs (all $ps < .001$). For the

interaction between test and SOA, follow-up t tests revealed a greater RT reduction after training at the three longer SOAs than at the first SOA (all $ps < .001$). Follow-up t tests for the interaction between compatibility and SOA revealed larger compatibility effects for the two longer SOAs than for the two shorter SOAs (all $ps < .003$).

A significant three-way interaction emerged between training, test, and compatibility; follow-up t tests revealed that compatibility effects increased after mirror training ($p = .002$) but did not change after countermirror training ($p = .177$). This three-way interaction was further modulated by SOA, as suggested by the significant four-way interaction between training, test, compatibility, and SOA. Follow-up t tests of the four-way interaction (Fig. 2, Table 2) revealed that automatic imitation increased by 19 ms from 42 ms at each of the two longer SOAs after mirror training [920 ms, $t(30) = 3.25$; 1,104 ms, $t(30) = 3.03$; all $ps < .006$], and decreased by 16 ms from 49 ms at the longest SOA after countermirror training [1,104 ms, $t(30) = 2.07$, $p = .047$].

Discussion

In this study, we investigated sensorimotor training effects on the automatic imitation of visual speech. We found that automatic imitation increased after mirror training and decreased after countermirror training. Moreover, mirror training had stronger effects than countermirror training. Our findings are largely consistent with the ASL hypothesis that the 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 excitatory matching links, which consequently enhanced observation-induced motor activation, leading to more facilitation in the compatible than in the incompatible condition (i.e., increased automatic imitation). ASL also proposes that sensorimotor experience of observing and executing different actions leads to excitatory nonmatching links between the sensory and motor representations of different actions, and also establishes inhibitory matching links between the sensory and motor representations of the same actions (Heyes et al., 2005). Accordingly, countermirror training in this study established inhibitory matching links that consequently reduced observation-induced motor activation, leading to less facilitation in the compatible than in the incompatible condition (i.e., decreased automatic imitation).

Fig. 2 **a** Mean response times $\pm$ standard errors in each experimental condition. Four panels represent pretraining (top left) and posttraining (top right) sessions in the mirror group and the pretraining (bottom left) and posttraining (bottom right) sessions in the countermirror group. **b**

Automatic-imitation effects (i.e., incompatible – compatible) $\pm$ standard errors for the pretraining (gray) and posttraining (black) sessions at each stimulus onset asynchrony in each group. *Significant changes in automatic-imitation effects after training ($p < .05$)

Training effects were found only for longer SOAs, where automatic imitation was greater. Speech actions consist of sequences of movements, and SOAs have been included in speech SRC tasks in order to demonstrate the time course of automatic imitation influenced by the different movement components of perceived speech. Potentially, the greater automatic imitation at longer SOAs was elicited by perceptually more salient components of perceived actions, and participants also paid more attention to these components during

training. Consequently, the automatic imitation elicited by these components was more susceptible to training. Future research could manipulate participants' attention to different action components during training and thus examine whether such manipulations influence training effects.

Importantly, automatic imitation of speech seems more resilient to countermirror training than is automatic imitation of manual actions. In Heyes et al. (2005), automatic imitation of manual movements was eliminated following countermirror

Table 1 Four-way ANOVA summary for log-transformed response times, as a function of training, test, compatibility, and stimulus onset asynchrony (SOA)

Main Effect/Interaction	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
Training	1, 60	0.41	.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 $\times$ Test	1, 60	2.53	.117	.041
Training $\times$ Compatibility	1, 60	0.26	.609	.004
Training $\times$ SOA	1.68, 100.52	0.68	.486	.011
Test $\times$ Compatibility	1, 60	3.56	.064	.056
Test $\times$ SOA	2.59, 155.35	9.67	< .001	.139
Compatibility $\times$ SOA	2.31, 138.81	18.92	< .001	.240
Training $\times$ Test $\times$ Compatibility	1, 60	12.99	.001	.178
Training $\times$ Test $\times$ SOA	2.59, 155.35	0.04	.984	.001
Training $\times$ Compatibility $\times$ SOA	2.31, 138.81	1.61	.199	.026
Test $\times$ Compatibility $\times$ SOA	3, 180	0.85	.465	.014
Training $\times$ Test $\times$ Compatibility $\times$ SOA	3, 180	3.42	.019	.054

Highlighted in bold = significant effects/interactions with $p < .05$

Table 2 Mean log-transformed response times, back-transformed response times in milliseconds, standard errors (SEs), and 95% confidence intervals (CIs) for each experimental condition in the testing sessions

Group			SOA	<i>M</i>	RT (ms)	<i>SE</i>	95% CI
Mirror	Pretraining	Compatible	1	6.355	575	0.030	[6.296, 6.414]
			2	6.314	552	0.030	[6.254, 6.374]
			3	6.262	524	0.031	[6.201, 6.323]
			4	6.232	509	0.027	[6.179, 6.286]
		Incompatible	1	6.401	602	0.029	[6.343, 6.458]
			2	6.363	580	0.026	[6.312, 6.414]
			3	6.339	566	0.025	[6.290, 6.388]
			4	6.311	551	0.023	[6.265, 6.357]
	Posttraining	Compatible	1	6.272	530	0.020	[6.233, 6.312]
			2	6.200	493	0.021	[6.159, 6.241]
			3	6.131	460	0.022	[6.087, 6.175]
			4	6.104	448	0.020	[6.064, 6.145]
		Incompatible	1	6.326	559	0.020	[6.286, 6.366]
			2	6.280	534	0.018	[6.243, 6.317]
			3	6.256	521	0.019	[6.217, 6.295]
			4	6.231	508	0.020	[6.191, 6.272]
Counter-Mirror	Pretraining	Compatible	1	6.318	554	0.030	[6.258, 6.377]
			2	6.261	524	0.030	[6.201, 6.321]
			3	6.205	495	0.031	[6.144, 6.266]
			4	6.186	486	0.027	[6.133, 6.240]
		Incompatible	1	6.368	583	0.029	[6.310, 6.425]
			2	6.335	564	0.026	[6.284, 6.386]
			3	6.301	545	0.025	[6.251, 6.350]
			4	6.282	535	0.023	[6.236, 6.328]
	Posttraining	Compatible	1	6.286	537	0.020	[6.246, 6.326]
			2	6.206	496	0.021	[6.165, 6.248]
			3	6.145	466	0.022	[6.102, 6.189]
			4	6.138	463	0.020	[6.097, 6.178]
		Incompatible	1	6.332	562	0.020	[6.292, 6.372]
			2	6.280	534	0.018	[6.243, 6.316]
			3	6.230	508	0.019	[6.191, 6.269]
			4	6.207	496	0.020	[6.167, 6.247]

SOA = stimulus onset asynchrony; SOA 1 = 552 ms; SOA 2 = 736 ms; SOA 3 = 920 ms; SOA 4 = 1,104 ms; RT = back-transformed response time

training. However, the automatic imitation of speech actions in our study was only reduced after countermirror training (960 trials in total) that was considerably longer than the training implemented by Heyes et al. (432 trials). Following ASL, this result may be accounted for by sociocultural imitative experience; it is possible that sensorimotor experience of observing and executing the same orofacial movements is gained mostly through social interactions. In monkeys, mouth mirror neurons have been found to be connected to the brain regions involved in emotion/reward processing, which plays a role in social activities (Ferrari, Gerbella, Coudé, & Rozzi, 2017). Comparable mirror activation has also been found in homologous regions in humans during the perception and

production of emotional facial expressions (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003). Hence, whereas both manual and orofacial observation-execution links are likely to result from sensorimotor learning, the extent of social influence may differ between the two, with the latter requiring more social engagement. Future studies could investigate how social manipulations modulate sensorimotor training effects on the automatic imitation of manual and orofacial movements.

Mirror training was more effective than countermirror training in our study, showing results opposite to those previously reported (Cracco et al., 2018). The ASL hypothesis proposes that observation-execution links could also be built

through a common stimulus (e.g., hearing people say /ba/) that co-occurs with both experience of performing a movement (e.g., saying /ba/) and experience of seeing others perform the same movement (e.g., seeing people say /ba/; Heyes, 2005). Such sensorimotor experience may contribute to the creation of indirect observation-execution links different from the direct ones learned through the training provided in our study. Thus, direct and indirect links might exist that are acquired through different experience, and the latter could have initially been stronger for speech actions, which are inherently multimodal. Therefore, our finding was likely due to initially weak direct links that were more susceptible to mirror than to counter-mirror training in modulating the automatic imitation of speech.

Our results do not allow us to completely exclude the possibility of an innate mechanism governing imitative, rather than counter- or nonimitative, sensorimotor associations underlying the imitation of orofacial movements. Heyes (2011) also acknowledged that results from training studies in principle “do not exclude a role for genetic prespecification in establishing the long-term sensorimotor connections that generate automatic imitation” (p. 478). Nevertheless, though the present study does not conclusively support ASL, our results are in line with the main ASL hypothesis that it is sensorimotor experience, not either sensory or motor experience alone, that configures observation-execution links (Heyes, 2010), since the only difference between the two groups in our study was the relationship between the observed and executed movements per trial during training. Hence, our findings suggest that the ASL mechanism can also be applied to communicative orofacial movements that infants learn to perceive and produce in the first few years of life. The precise mechanisms responsible for forging the sensorimotor associations underlying speech actions could be further explored by providing extended counter-mirror training. If extended training leads to a reduction/reversal of automatic imitation, this would support the notion that learning of these associations is not necessarily constrained by innate factors.

The simulation theory of speech perception proposes that observation-induced motor activation facilitates prediction of the incoming signals supporting speech comprehension (Pickering & Garrod, 2013). Critically, greater motor involvement is suggested when observers have more experience with the perceived speech. By applying transcranial magnetic stimulation (TMS) to lip motor cortex (lip M1), Swaminathan et al. (2013) found facilitated lip M1 excitability during the viewing of sentences spoken in a known as compared to an unknown language. Following ASL, Swaminathan et al. suggested that the difference between the two conditions was due to different strengths of perception–production links underlying known and unknown languages, hence supporting the simulation hypothesis that more experience leads to greater observation-induced motor activation. Our results

further suggest that it was imitative sensorimotor learning that facilitated the observation-induced motor activation. Additionally, overt imitation of accented speech improves subsequent speech perception, indicating that imitative learning leads to enhanced observation-induced motor activation facilitating speech comprehension (Adank, Hagoort, & Bekkering, 2010). Schmitz et al. (2018) stimulated lip M1 with TMS and found that listening to nonnative vowels elicited higher articulatory excitability than did native-like vowels, which was opposite to what Swaminathan et al. found with sentence articulations presented visually. Future research that controls linguistic levels and stimulus modalities will be required in order to investigate this inconsistency. Moreover, follow-up research could extend our findings by examining whether sensorimotor training modulates the audio–motor links underlying speech. Behavioral research could also examine automatic imitation of nonnative speech and investigate the role of sensorimotor learning in establishing perception–production links underlying second language processing.

In conclusion, the present study showed that sensorimotor training modulated automatic imitation of visual speech. As such, our results elucidate questions concerning the flexibility of imitative mechanisms and add to the growing body of evidence on perception–production links in speech processing.

Author note This work was supported by a grant of the China Scholarship Council to the first author and by the BIAL Foundation under Grant No. 267/14 to P.A.

Open practices statement The data and materials for this experiment are available upon request, and no experiment was preregistered.

References

- Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, 21, 1903–1909.
- Adank, P., Nuttall, H., Bekkering, H., & Maegherman, G. (2018). Effects of stimulus response compatibility on covert imitation of vowels. *Attention, Perception, & Psychophysics*, 80, 1290–1299. <https://doi.org/10.3758/s13414-018-1501-3>
- Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Retrieved from www.praat.org
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., & Rizzolatti, G. (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron*, 42, 323–334.
- Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences*, 100, 5497–5502.
- Casile, A., Caggiano, V., & Ferrari, P. F. (2011). The mirror neuron system: A fresh view. *Neuroscientist*, 17, 524–538.
- Catmur, C. (2013). Sensorimotor learning and the ontogeny of the mirror neuron system. *Neuroscience Letters*, 540, 21–27.

- Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: The role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364, 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, 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, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- 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, 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, 15061–15065.
- Heyes, C. (2005). Imitation by association. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science. Vol. 1: Mechanisms of imitation and imitation in animals* (pp. 157–176). Cambridge, MA: MIT Press.
- Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, 34, 575–583.
- Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, 137, 463–483. <https://doi.org/10.1037/a0022288>
- Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research*, 22, 233–240.
- Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 634–647. <https://doi.org/10.1037/0096-1523.26.2.634>
- Meltzoff, A. N. (2002). Elements of a developmental theory of imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind development, evolution and brain bases* (pp. 19–41). Cambridge, UK: Cambridge University Press.
- Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience and Biobehavioral Reviews*, 36, 341–349.
- Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, 20, 750–756.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329–347.
- Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society B*, 274, 2509–2514.
- Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science*, 14, 92–105.
- Schmitz, J., Bartoli, E., Maffongelli, L., Fadiga, L., Sebastian-Galles, N., & D'Ausilio, A. (2018). Motor cortex compensates for lack of sensory and motor experience during auditory speech perception. *Neuropsychologia*. Advance online publication. <https://doi.org/10.1016/j.neuropsychologia.2018.01.006>
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105.
- Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1746–1759. <https://doi.org/10.1037/0096-1523.26.6.1746>
- Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., & Möttönen, R. (2013). Motor excitability during visual perception of known and unknown spoken languages. *Brain and Language*, 126, 1–7. <https://doi.org/10.1016/j.bandl.2013.03.002>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.