

Interaction in planning movement direction for articulatory gestures and manual actions

Lari Vainio¹ · Mikko Tiainen¹ · Kaisa Tiippana¹ · Naeem Komeilipoor¹ · Martti Vainio²

Received: 17 April 2015 / Accepted: 18 June 2015 / Published online: 1 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Some theories concerning speech mechanisms assume that overlapping representations are involved in programming certain articulatory gestures and hand actions. The present study investigated whether planning of movement direction for articulatory gestures and manual actions could interact. The participants were presented with written vowels (Experiment 1) or syllables (Experiment 2) that were associated with forward or backward movement of tongue (e.g., [i] vs. [a] or [te] vs. [ke], respectively). They were required to pronounce the speech unit and simultaneously move the joystick forward or backward according to the color of the stimulus. Manual and vocal responses were performed relatively rapidly when the articulation and the hand action required movement into the same direction. The study suggests that planning horizontal tongue movements for articulation shares overlapping neural mechanisms with planning horizontal movement direction of hand actions.

Keywords Speech · Hand movement · Articulatory gesture · Movement direction · Motor priming

Introduction

Long-standing theories concerning integration between hand and mouth movements (i.e., so-called mouth-gesture theories) assume that vocal language, or at least some articulatory gestures of speech, may have evolved from manual gestures. For example, according to Paget (1930) and Hewes (1973), people may have an innate tendency to unintentionally copy their hand gestures by movements of tongue and lips. Similar tendency has been also observed in captive chimpanzees (Waters and Fouts 2002). As a consequence, some articulatory gestures can be assumed to be evolved, at least to some extent, from mechanisms underlying manual gestures. Hence, it can be speculated that overlapping representations are involved in planning certain articulatory gestures and hand movements. The present study asks whether overlapping representations might be involved in planning the horizontal movements of hand and articulation (e.g., moving the hand forward and moving the tongue forward for pronouncing a front vowel).

Over the last few decades, a growing body of evidence has shown that common representations are indeed involved in programming hand and speech acts. For instance, overlapping brain networks appear to be involved in speech and hand actions (e.g., Binkofski and Buccino 2004; Iverson and Thelen 1999). In particular, this interplay between speech and hand movements has been linked to grasp actions. For example, one of the most influential gestural theories of speech evolution proposes that the evolution of grasping and an imitation system for grasping are the milestones on which the speech system has initially evolved (Arbib 2005). In line with this theory, in monkey, single-cell recordings (Rizzolatti and Gentilucci 1988; Gentilucci and Rizzolatti 1990) and electrical stimulation studies (Graziano and Aflalo 2007) have demonstrated that

✉ Lari Vainio
lari.vainio@helsinki.fi

¹ Division of Cognitive and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, Siltavuorenpenger 5 A, PL 9, 00014 Helsinki, Finland

² Phonetics and Speech Synthesis Research Group, Institute of Behavioural Sciences, University of Helsinki, Siltavuorenpenger 5 A, PL 9, 00014 Helsinki, Finland

the same neurons are involved in commanding manual grasp motor acts and mouth motor acts in the premotor area F5, which is considered the homologue of human Broca's area (Rizzolatti et al. 1988). Furthermore, several behavioral studies have shown integration between hand grasping and articulatory mouth movements in human (Gentilucci et al. 2001; Vainio et al. 2014). For example, a spontaneous increase in the opening of a grasp shape has been reported when the behavioral task requires simultaneous articulation of an open vowel [a] in comparison with articulating a close vowel [i] (Gentilucci and Campione 2011).

The behavioral evidence that is particularly in line with the idea of mouth-gesture theories, according to which some articulatory gestures may be oral versions of hand actions, is provided by our recent study (Vainio et al. 2013). We have demonstrated systematic connections between planning precision and power grip actions and certain articulatory gestures. We found that the participants performed precision grip responses faster and more accurately than power grip responses if they were simultaneously articulating a close vowel [i] or the voiceless stop consonant [t]. In contrast, similar improvement in power grip responses was observed if the participants were articulating an open vowel [a] or the voiceless stop consonant [k]. In other words, articulatory gestures in which the tip of the tongue is brought into contact with the alveolar ridge and the teeth, or in which the aperture of the vocal tract remains relatively small, might be, to some extent, oral versions of the precision grip and consequently might be planned in overlapping representations with the precision grip. In contrast, the articulatory gestures that are produced by moving the back of the tongue against the velum, or in which the aperture of the vocal tract remains relatively large, might be, to some extent, oral versions of the power grip.

The above-mentioned effect between the grip type and articulatory gesture was proposed to support the view that there is anatomical, functional, and evolutionary associations between speech and hand gestures. On the evolutionary scale, speech might have not been the initial function of this connection between mouth and hand movements. Rather it is more likely that the manual actions were first connected to mouth actions, for example, for ingestive purposes (Gentilucci et al. 2001), and gradually these connections were also adapted for speech purposes, such as for shaping articulatory gestures.

Importantly for the purpose of the current study, it has been speculated that, in addition to the already-mentioned interplay between hand grasping and articulatory gestures, horizontal hand movements (i.e., moving the hand away from the body or toward the body) might also be programmed in common representations with mouth movements. Ramachandran and Hubbard (2001) proposed that articulatory gestures in some words such as 'you' might

mimic pointing forward as the articulation of that word requires pouting ones lips forward. In contrast, articulatory gestures in some other words such as 'I' might mimic pointing backward toward oneself. These kinds of examples could be assumed to be hand-related versions of sound symbolism phenomena (Sapir 1929) in which articulatory gestures related to production of some words convey information about the actual meaning of these words (e.g., a tendency to use words containing the vowel [i] for smaller objects and words containing the vowels [a] and [o] for larger objects) (Ramachandran and Hubbard 2001). However, in contrast to typical sound symbolism phenomena, in these hand-related versions, the articulatory gesture is echoing corresponding manual gesture rather than, for example, size or shape properties of objects.

Some studies have shown that horizontal hand movements can be automatically associated with the semantic content of certain words. Glenberg and Kaschak (2002) famously showed that participants are faster in moving their hand away from their body if they are simultaneously making judgments about whether the sentence, which implies action away from the body (e.g., 'Close the drawer'), is sensible. In contrast, hand movements toward their body were facilitated if the sentence implied action toward the body (e.g., 'Open the drawer'). Moreover, Chieffi et al. (2009) showed that the hand movements toward the body are performed faster if participants are simultaneously reading the word 'QUA' ('here'), and the hand movements away from the body are performed faster if participants are reading the word 'LA' ('there'). However, regardless of the intuitively plausible interplay between horizontal hand and mouth movements, to our knowledge, the idea has not been previously investigated at the level of articulatory gestures. The study that comes closest to exploring this idea empirically has been carried out by Higginbotham et al. (2008). They showed that manual pointing movements as well as precision grasping can be associated with automatic and spontaneous increase in electromyographic responses of the orbicularis oris muscles—the muscles that are required to produce lip movements in which lips are moved forward in order to, for example, produce certain vowels such as [y] or [o].

The current study investigates whether interaction between horizontal hand movements and speech can operate at the level of articulatory gestures, in particular, in relation to articulatory tongue movements. Although this idea has not been previously researched, it is highly plausible to assume that planning the direction of hand movements and articulatory gestures would interact at some level of processing. It has been, for example, found that two hands are preferably moved in the same direction than to opposite directions (Serrien et al. 1999), suggesting that preparation of movement direction of different effectors

shares the same directional code at some level of the action preparation. Given that the systems that plan hand and mouth actions are tightly connected, similar shared directional coding of mouth and hand movements could be predicted.

Experiments 1 and 2 investigated whether spatial compatibility in movement direction between overt articulation and horizontal hand movement leads to a congruency effect in manual and vocal responses. We employed a modified version of the speeded choice reaction task originally used by Vainio et al. (2013). In the current study, the participants were presented with meaningless speech units (e.g., [i]) written on a computer screen. The participants were required to move a joystick either forward or backward according to the color (green or blue) of the text and simultaneously pronounce the speech unit. We measured reaction times and accuracy of manual and vocal responses.

In Experiment 1, the participants were required to pronounce a vowel while they performed either a push or pull action with their hand. The vowels that we used were [i], [a], and [o]. In Finnish, the vowel [i] is the most frontal vowel. In contrast, the vowels [a] and [o] are both back vowels from which the former is a back unrounded vowel, whereas the latter is a back rounded vowel. Although it is generally assumed that the vowel [a] is slightly more back vowel than the vowel [o], our own preliminary pilot data suggested that in fact, [o] might be even more back vowel than [a] in Finnish—in terms of F2 (formant 2) values known to reflect the frontness of tongue movement in articulation (Fant 1960). Consequently, both of these back vowels were selected for Experiment 1. If the assumption concerning the interaction between horizontal hand movements and articulation is correct, we should observe a congruency effect in relation to manual and vocal responses when the utterance requires articulatory movements into the same direction as the hand movement (e.g., [i]/forward).

Experiment 2 investigated whether articulation of the consonants [t] or [k] during a horizontal hand movement would reveal shared direction coding between tongue-based articulation and hand movement. The consonant [t] is associated with horizontal forward movement of a tongue, whereas the consonant [k] is associated with horizontal backward movement of a tongue. In more detail, [t] is produced by bringing the tip of the tongue into contact with the alveolar ridge and the teeth. In contrast, [k] is produced by moving the back of the tongue up against the velum. The consonant [k] was the obvious choice for back consonant as it is the only commonly used back consonant in Finnish. Regarding the front consonant, [t] was selected because it is the most front consonant and because it is a voiceless stop consonant like [k]. The consonants were coupled with the vowel [e] because it is difficult to pronounce them alone. The vowel [e] was selected because it is relatively neutral

in terms of tongue frontness at least in comparison with the vowels [i], [a], and [o].

In addition to measuring reaction times and accuracy of manual and vocal responses, we also anticipated that the compatibility or incompatibility between movements of hand and articulation might have some influence on certain elements of the voice spectrum. Hence, we measured the following acoustic features of the vocalizations: intensity, f_0 (fundamental frequency), F1 (formant 1), and F2. For instance, it might be speculated that manual push movement could increase F2 values. That is, because F2 is known to increase as a function of forward movement of the tongue in articulation (Fant 1960), manual push movement might lead to a slight increase in simultaneous forward tongue movement.

Experiment 1

Methods

Participants and ethical review

Twenty-two naïve volunteers participated in Experiment 1 (21–41 years of age; mean age = 29.7 years; 6 males). All participants were native speakers of Finnish and had normal or corrected-to-normal vision and were right-handed. Written informed consent was obtained from all participants. The study was approved by the Ethical Committee of the Institute of Behavioural Sciences at the University of Helsinki and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, stimuli and procedure

Each participant sat in a dimly lit room with his or her head 65 cm in front of a 19-in. LCD monitor (screen refresh rate: 75 Hz; screen resolution: 1280 × 1024). The head-mounted microphone was adjusted close to the participant's mouth for recording vocal responses. The manual response device was a joystick that was located between the participant and the monitor, 30 cm away from the monitor. The joystick was attached steadily onto the response table. The joystick was vertically positioned so that the up-most part of the stick was at the level of the low-most part of the monitor. It was positioned horizontally at the center of the monitor.

In total, the experiment consisted of 240 trials [30 (repetition) × 2 (block) × 2 (vowel) × 2 (direction)]. The experiment began with practice trials. Each participant was given as much practice as it took to perform the task fluently. The experiment consisted of two blocks that were separated by a short break. The participants were instructed to continue

from the break by pressing the space bar of a keyboard that was positioned between the joystick and the monitor. Both blocks consisted of two different target vowels (Block 1: [i] and [a]; Block 2: [i] and [o]) that were presented at the center of the monitor. The vowels were written in KaiTi font (lowercase; bold; font size: 72). The vowels were presented in randomized order within both blocks. In addition, the order of blocks was balanced between the participants.

Participants held the joystick in their right hand. At the beginning of each trial, a fixation cross was presented for 400 ms at the center of the screen. Then, a blank screen was displayed for 400 ms. After that the target vowel was presented at the center of the screen and remained in view for 1000 ms or until a response was made. Finally, a blank screen was displayed for 500 ms. The target was presented either in green or in blue color. The participants' task was to pull or push the joystick until the end of the full motion range of the joystick (4.5 cm forward or backward) as fast and accurately as possible according to the color of the stimulus. Half of the participants responded to the green by pulling the joystick, and other half responded to the blue by pulling the joystick. The response directions were marked with corresponding color tapes on the response device. The joystick was always returned to the central starting position after the response. The mechanisms of the joystick provided a minor force that pulled the stick back toward the starting position. All stimuli were displayed on a gray background. Erroneous manual responses were immediately followed by a short 'beep' tone.

In addition to the manual response, the participants were instructed to pronounce the presented vowel as fast as possible. It was emphasized that the vowel should be uttered in natural talking voice at the same time with the manual response. The actual experiment was not started before the participant demonstrated in the practice session that he/she continuously produced the vocal and manual reactions at the same time. The recording levels of the vocal responses were calibrated individually for each subject at the beginning of the experiment. Stimulus presentation and sound recording were done with Presentation 16.1 software (Neurobehavioral Systems Inc., Albany, CA, USA).

Vocal reaction times were measured for 1500 ms from the onset of the target object to the onset of the vocalization. Onsets of the vocalizations were located individually for each trial using Praat (v. 5.3.49) (<http://www.praat.org>). Manual reaction times (RTs) were measured from the onset of the target stimulus to the point of the joystick movement in which the joystick exceeded 10 % from the motion range of the joystick. Only those manual reactions in which the movement was performed completely (i.e., until the end of the full motion range of the joystick) were accepted to the RT analysis. Moreover, the response was registered as an 'error' if joystick movement exceeded 50 % of the motion

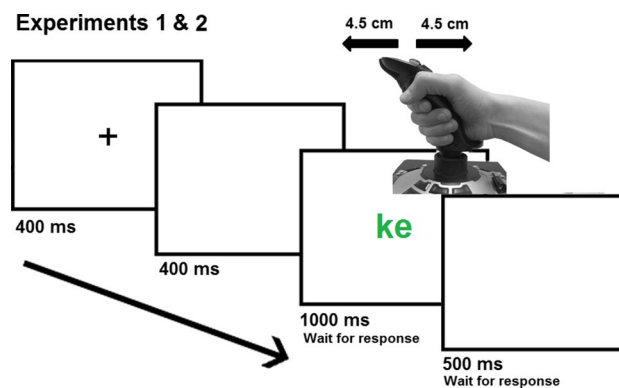


Fig. 1 Graphical presentation of the trial structure of Experiments 1 and 2. See text for details

range in the wrong direction. In addition, the response was registered as 'no-response' if the joystick movement did not exceed 50 % of the motion range (regardless of the correctness of its direction) or the movement was performed to the correct direction, and it exceeded 50 % of the motion range, but it did not exceed 100 % of the motion range (Fig. 1).

Results and discussion

In total, 9.5 % of the raw data were discarded from the analysis of *manual reaction times* including .5 % of trials containing vocal errors, 3.8 % of trials containing manual errors, .3 % of trials containing no-responses, and 4.9 % of trials in which the manual RTs were more than two standard deviations from a participant's overall mean. In total, 8.5 % of the raw data were discarded from the analysis of *vocal reaction times* including .5 % of trials containing vocal errors, 3.8 % of trials containing manual errors, .3 % of trials containing no-responses, and 3.9 % of trials in which the vocal RTs were more than two standard deviations from a participant's overall mean. Condition means for the remaining data were subjected to a repeated-measures ANOVA with the within-participants factors of block ([i]/[a] or [i]/[o]), vowel ([i] or [a]/[o]), and manual response (pull or push, i.e., backward or forward, respectively). Post hoc comparisons were performed by means of *t* tests applying a Bonferroni correction when appropriate. A partial eta-squared statistic served as effect size estimate.

The analysis of manual reaction times revealed a significant main effect of manual response. The participants pushed the joystick faster ($M = 413$ ms) than pulled it ($M = 427$ ms), $F_{(1,21)} = 19.98$, $p < .001$, $n_p^2 = .487$. More interestingly, the analysis revealed a significant interaction between vowel and manual response, $F_{(1,21)} = 24.73$, $p < .001$, $n_p^2 = .541$. This interaction is presented in Fig. 2. The pairwise comparisons test showed that the pull

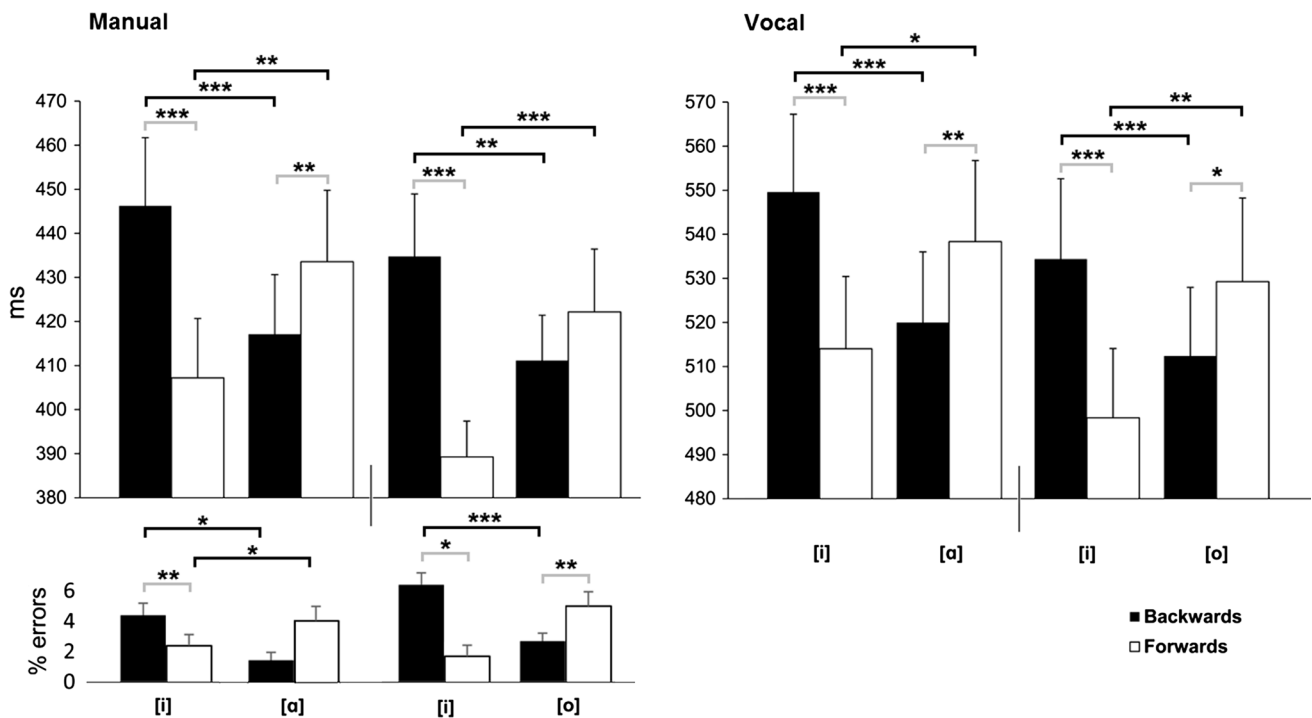


Fig. 2 Mean vocal reaction times and manual reaction times and errors in Experiment 1 as a function of the vowel and the movement direction. Manual and vocal responses were performed faster when the articulation and the hand action required movement into the same direction (e.g., forward manual response and the vowel [i]). The means for percentage error rates of manual responses are presented below the corresponding reaction time histogram. Error bars depict the standard error of the mean. Asterisks indicate statistically significant differences ($***p < .001$; $**p < .01$; $*p < .05$). In manual reaction times, the black horizontal brackets indicate that the reaction time associated with one type of manual movement (e.g., pull) was

significantly modulated by the pronounced vowel (e.g., [a] vs. [i]). The gray horizontal brackets indicate that there was a significant difference between push and pull reaction times when participants were pronouncing one type of vowel (e.g., [i]). In the vocal reaction times, the black horizontal brackets indicate that the movement direction of the hand (e.g., backward) caused that one vowel (e.g., [a]) was produced significantly faster than other vowel (e.g., [i]). The gray horizontal brackets indicate that a certain vowel (e.g., [i]) was produced significantly faster when the hand was moved to one direction (e.g., forward) rather than other direction (backward)

responses were performed significantly faster when the vowel was [a] ($M = 417$ ms) rather than [i] ($M = 446$ ms) ($p < .001$) in Block 1, and [o] ($M = 411$ ms) rather than [i] ($M = 435$ ms) ($p = .002$) in Block 2. In contrast, the push responses were made significantly faster, when the vowel was [i] ($M = 407$ ms) rather than [a] ($M = 434$ ms) ($p = .009$) in Block 1, and [i] ($M = 389$ ms) rather than [o] ($M = 422$ ms) ($p = .001$) in Block 2. Finally, the difference between push and pull reaction times was significant when participants were pronouncing [i] (pull: $M = 446$ ms; push: $M = 407$ ms; $p < .001$) in Block 1, [a] (pull: $M = 417$ ms; push: $M = 434$ ms; $p = .009$) in Block 1, and [i] (pull: $M = 435$ ms; push: $M = 389$ ms; $p < .001$) in Block 2.

An analysis of percentage error rates of manual responses revealed a significant interaction between vowel and manual response, $F_{(1,21)} = 16.53$, $p = .001$, $\eta_p^2 = .441$. This interaction is presented in Fig. 2. The pairwise comparisons test showed that the accuracy of manual responses was significantly improved by the compatibility between the response direction and the vowel (e.g., push [i])

when the participant pulled the joystick and simultaneously pronounced the vowel [a] ($M = 2.4$ %) rather than [i] ($M = 4.4$ %) ($p = .024$) in Block 1, or the vowel [o] ($M = 1.7$ %) rather than [i] ($M = 6.4$ %) ($p < .001$) in Block 2. In contrast, participants made significantly fewer errors when they had to push the joystick and pronounce the syllable [i] ($M = 1.5$ %) rather than [a] ($M = 4.0$ %) ($p = .013$) in Block 1. Finally, the difference between the error rates of push and pull responses was significant when participants were pronouncing [i] in Block 1 (pull: $M = 4.4$ %; push: $M = 1.5$ %; $p = .004$) and 2 (pull: $M = 6.4$ %; push: $M = 2.7$ %; $p = .026$) and when they were pronouncing the vowel [o] in Block 2 (pull: $M = 1.7$ %; push: $M = 5.0$ %; $p = .011$).

The analysis of vocal reaction times revealed a significant main effect of manual response. Vocalizations were performed faster when the participant was simultaneously making the push response ($M = 520$ ms) rather than the pull response ($M = 529$ ms), $F_{(1,21)} = 9.01$, $p = .007$, $\eta_p^2 = .300$. More interestingly, the analysis revealed a

significant interaction between syllable and manual response, $F_{(1,21)} = 21.13$, $p < .001$, $\eta_p^2 = .502$. This interaction is presented in Fig. 2. The pairwise comparisons test showed that, in Block 1, [i] responses were produced faster when the participants were simultaneously pushing the joystick ($M = 514$ ms) rather than pulling it ($M = 550$ ms) ($p < .001$). Similarly, in Block 2, [i] responses were also produced faster when the participants were simultaneously pushing the joystick ($M = 498$ ms) rather than pulling it ($M = 534$ ms) ($p < .001$). In contrast, in Block 1, [a] responses were produced faster when the participants were simultaneously pulling the joystick ($M = 520$ ms) rather than pushing it ($M = 538$ ms) ($p = .010$). Similarly, in Block 2, [o] responses were produced faster when the participants were simultaneously pulling the joystick ($M = 512$ ms) rather than pushing it ($M = 529$ ms) ($p = .045$). Finally, when the participants were pulling the joystick, they pronounced (in Block 1) the vowel [a] ($M = 520$ ms) faster than [i] ($M = 550$ ms) ($p < .001$) and in Block 2 the vowel [o] ($M = 512$ ms) faster than [i] ($M = 534$ ms) ($p = .001$). In contrast, when the participants were pushing the joystick, they pronounced the vowel [i] ($M = 514$ ms) faster than [a] ($M = 538$ ms) ($p = .028$) in Block 1 and the vowel [i] ($M = 498$ ms) faster than [o] ($M = 529$ ms) ($p = .002$) in Block 2. Because the participants produced only 38 vocal errors, we did not analyze the vocal errors.

Voice characteristics The spectral characteristics of the vocal data were analyzed using Praat. For the intensity value, the peak intensity of the voiced section was selected. The f_0 and formants F1 and F2 were all calculated as a median value from the middle third of the vocalization area. Then, the values more than two standard deviations from each participant's condition means of intensity, f_0 , F1 and F2 were excluded from the voice spectra analysis. In general, the manual performance was not observed to influence voice characteristics. The only consistent pattern of voice characteristics observed in F1 and F2 was the difference between the different vowels. This effect manifested itself in a significant interaction between block and vowel [F1: $F_{(1,21)} = 49.66$, $p < .001$, $\eta_p^2 = .703$; F2: $F_{(1,21)} = 13.96$, $p < .001$, $\eta_p^2 = .399$]. Regarding the spectral component of F1, the values were lower for the front vowels (Block 1—[i]: 356 Hz; Block 2—[i]: $M = 359$ Hz) than back vowels (Block 1—[a]: 627 Hz; Block 2—[o]: $M = 517$ Hz). In contrast, F2 was higher for the front vowels (Block 1—[i]: 2607 Hz; Block 2—[i]: $M = 2597$ Hz) than back vowels (Block 1—[a]: 1126 Hz; Block 2—[o]: $M = 975$ Hz).

In line with our hypothesis, the results of Experiment 1 showed clearly that manual push responses were performed relatively rapidly and accurately when the front vowel [i] was pronounced during the manual response, while manual pull responses were performed relatively rapidly and

accurately when the back vowels [a] and [o] were pronounced during the manual response. In addition, this congruency effect was also observed in vocal reaction times. When the direction of the hand movement and frontness/backness of the pronounced vowel were congruent, the vocal responses were produced relatively rapidly.

Experiment 2

Methods

Participants and ethical review

Sixteen naïve volunteers participated in Experiment 2 (20–29 years of age; mean age = 22.4 years; 3 males). All participants were native speakers of Finnish and had normal or corrected-to-normal vision and were right-handed. We obtained written informed consent from all participants. The study was approved by the Ethical Committee of the Institute of Behavioural Sciences at the University of Helsinki and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, stimuli, and procedure

The apparatus, environmental conditions, procedure, and calibration were the same as those in Experiment 1. Also, the stimuli arrangements were identical to those of Experiment 1 with the exception that the stimuli consisted of the syllables [te] and [ke]. Manual reaction times, vocal reaction times, errors, and no-responses were measured in the same way as in Experiment 1. In total, the experiment consisted of 120 trials [30 (repetition) \times 2 (syllable) \times 2 (direction)].

Results and discussion

In total, 10.8 % of the raw data were discarded from the analysis of *manual reaction times* including .6 % of trials containing vocal errors, 3.8 % of trials containing manual errors, 2.3 % of trials containing no-responses, and 4.1 % of trials in which the manual RTs were more than two standard deviations from a participant's overall mean. In total, 11.3 % of the raw data were discarded from the analysis of *vocal reaction times* including .6 % of trials containing vocal errors, 3.8 % of trials containing manual errors, 2.3 % of trials containing no-responses, and 4.6 % of trials in which the vocal RTs were more than two standard deviations from a participant's overall mean. Condition means for the remaining data were subjected to a repeated-measures ANOVA with the within-participants factors of

syllable ([te] or [ke]) and manual response (pull or push, i.e., backward or forward, respectively). Post hoc comparisons were performed by means of *t* tests applying a Bonferroni correction when appropriate. A partial eta-squared statistic served as effect size estimate.

The analysis of manual reaction times revealed a significant interaction between syllable and manual response, $F_{(1,15)} = 7.82$, $p = .014$, $\eta_p^2 = .343$. The push responses were performed marginally faster when the syllable was [te] rather than [ke] ($M = 427$ ms vs. $M = 448$ ms, $p = .055$). In contrast, the pull responses were performed faster when the syllable was [ke] rather than [te] ($M = 425$ ms vs. $M = 445$ ms, $p = .014$). Finally, the difference between push and pull reaction times was significant when participants were pronouncing [te] (push: $M = 427$ ms; pull: $M = 445$ ms; $p = .035$) and [ke] (push: $M = 448$ ms; pull: $M = 425$ ms; $p = .045$). In contrast to the results of Experiment 1, an analysis of percentage error rates of manual responses did not reveal any significant main effects or interactions. This is not surprising since each participant made on average only five errors out of 120 trials (i.e., the mean error rate was 4 %). Thus, in Experiment 2, the experimental power was not adequate to reveal any significant effect in the error analysis.

The analysis of vocal reaction times revealed a significant interaction between syllable and manual response, $F_{(1,15)} = 7.16$, $p = .017$, $\eta_p^2 = .323$. [te] responses were produced faster when the participants were simultaneously pushing the joystick rather than pulling it ($M = 576$ vs. $M = 590$ ms, $p = .027$), and [ke] responses were produced faster when the participants were simultaneously pulling the joystick rather than pushing it ($M = 562$ ms vs. $M = 588$ ms, $p = .045$). Finally, when the participants were pulling the joystick, they pronounced the syllable [ke]

($M = 562$ ms) faster than [te] ($M = 590$ ms) ($p = .006$). An opposite pattern was observed when the participants were pushing the joystick even though this effect was not significant (syllable [te]: $M = 576$ ms; syllable [ke]: $M = 588$ ms; $p = .115$). Because the participants produced only 11 vocal errors, we did not analyze the vocal errors.

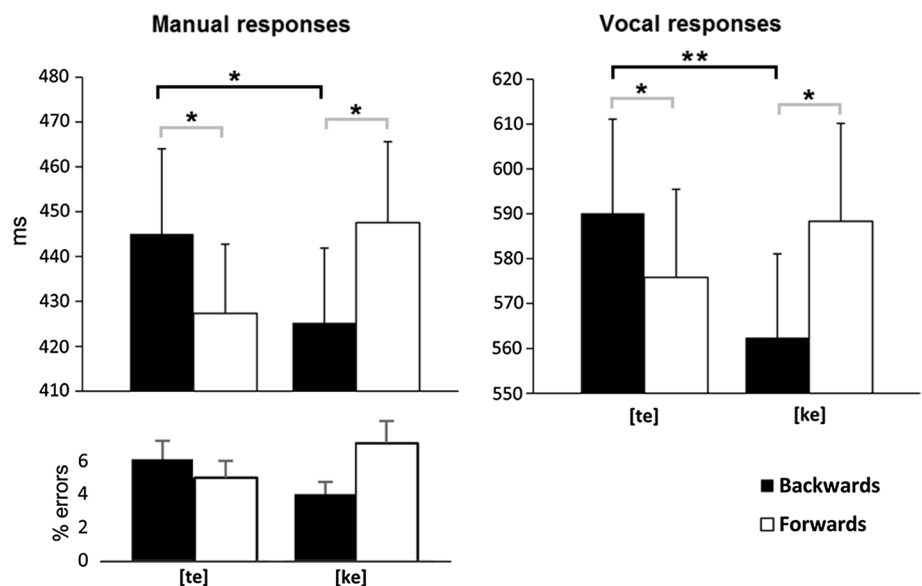
Voice characteristics The analysis of voice characteristics was carried out in the same way as in Experiment 1. That is, the analysis was carried out for middle third of the vowel section (i.e., for the vowel [e]) of the vocalization. Similar to the results of Experiment 1, the manual performance did not influence voice characteristics. The analysis of intensity revealed a significant main effect of syllable, $F_{(1,15)} = 56.27$, $p < .001$, $\eta_p^2 = .790$. The intensity was higher for [te] responses ($M = 82.5$ dB) than for [ke] responses ($M = 81.8$ dB). The analyses of f_0 and F2 did not reveal any significant main effects or interactions. This is not very surprising as in both analyzed syllables, the actual analyzed component was the same (i.e., the vowel [e]). However, the analysis of F1 revealed significant main effects of syllable [$F_{(1,15)} = 24.05$, $p < .001$, $\eta_p^2 = .616$]. F1 was higher for the syllable [te] ($M = 585$ Hz) than [ke] ($M = 573$ Hz).

The results of Experiment 2 revealed that the congruency effect observed in Experiment 1 with vowels can be also observed when participant has to select between two opposing consonants for the response (Fig. 3).

General discussion

Overall, our results suggest that direction codes of manual motor programs are implicitly planned, at least to some extent, in the same representations as the direction codes

Fig. 3 Mean vocal reaction times and manual reaction times and errors in Experiment 2 as a function of the syllable ([te] and [ke]) and the movement direction. Manual and vocal responses are performed faster when the articulation and the hand action required movement into the same direction. Other details are as in Fig. 2



of articulatory gestures. The study shows that manual and vocal responses were relatively fast when the articulation organs were moved in the same horizontal direction as the simultaneously performed manual response. In more detail, the congruency effect was observed with the front vowel [i] and the back vowels [a] and [o]. The front vowel was associated with push (i.e., forward) responses, and the back vowels were associated with pull (i.e., backward) responses. This effect was also observed with consonants. The front consonant [t] was associated with push responses, and the back consonant [k] was associated with pull responses. This outcome is consistent with the mouth-gesture theories according to which overlapping representations might be involved in planning certain articulatory gestures and hand movements. The finding is a novel addition to the current literature concerning integration between manual and articulatory planning processes as it suggests that the interplay between these two motor systems does not only operate at the level of grasp programming—a coupling that has been shown already (Gentilucci et al. 2001; Vainio et al. 2013; Gentilucci 2003)—but can also operate at the level of programming horizontal movements of articulatory gestures and manual actions.

We assume that the current findings reflect shared planning processes between articulation and hand movements. In everyday life, we are continuously under the pressure of multiple conflicting opportunities for action. The sensorimotor system exploits any potentially relevant information in order to decide the course of action and then prepare and execute the chosen motor plan (Cisek and Kalaska 2010). As an example, it would be possible that manual planning processes are biased by simultaneously planned articulatory processes if there would exist some functional connection between the two effectors. As such, we propose that the manual congruency effect observed in the present study is a result of competition between two potential actions required for manual responses (i.e., push vs. pull), which is biased by simultaneously occurring direction code processing for articulatory responses (i.e., whether to produce front or back vowel). The push or pull response is improved if there is a congruency in direction codes of speech-related and manual processes. This is to say that speech processes and manual processes share to some extent a common representation codes, at some level of the action preparation (cognitive, sensorimotor, or primary motor), for preparing the movement in a certain direction.

However, it remains unclear whether this direction code is shared only between the mouth and hand motor processes or whether this direction code operates in more general manner interacting with any kinds of motor (e.g., leg movements) or even cognitive functions (e.g., understanding semantic content for direction words such as ‘forward’). This issue warrants for further investigation. In

light of the current findings, we can only be sure that processes that prepare horizontal hand movements share the direction code with the processes that prepare articulatory gestures. Furthermore, this direction code appears to be shared between these two effectors in an implicit manner, firstly, because the participants were not required to explicitly process the frontness/backness of the speech units that they had to pronounce. Secondly, after the experiment, all of the participants reported that they were entirely unaware about the objectives of the experiment. They seemed genuinely surprised when the experimenter told them after the study that, for example, [i] is more frontal vowel than [o].

The other issue that needs to be considered is whether the observed interaction in preparing mouth and hand actions operates exclusively for speech-related functions or whether any push–pull mouth movements would interact with corresponding hand movements regardless of the purpose for which the mouth action is performed. However, as we already stated in the Introduction, our view is that, in the evolutionary scale, manual actions were initially connected to mouth actions for ingestive purposes, and gradually, these connections were also adapted for speech purposes, such as for shaping articulatory gestures. That is, our view is that the direction codes between mouth and hand movements are interacting at very basic and abstract level, and this common directional coding can be also—observed in speech-related oral tasks. We propose that people’s innate tendency to unintentionally copy their hand gestures by movements of tongue (Paget 1930; Hewes 1973) originates from these basic shared mechanisms in planning of goal states of tongue and hand with respect to, for example, movement direction. This in turn might have resulted in hand-related versions of sound symbolism phenomena in which the articulatory gesture is echoing corresponding manual gesture such as a pointing gesture (Ramachandran and Hubbard 2001). Nevertheless, further experimentation is needed to investigate whether the effect is exclusively associated with speech processes.

Finally, it is interesting to speculate potential connections between forward and backward hand movements and precision and power grasps. Similarly associating the vowel [i] and the consonant [t] with forward movements and the vowel [a] and the consonant [k] with backward movements, these same speech units have been also associated with precision grasping and power grasping, respectively (Vainio et al. 2013). It has been suggested that pointing movement is developed at the end of the first year through modeling out of grasp movements (Leung and Rheingold 1981; Murphy and Messer 1977). In line with this idea, it would be plausible to assume that programming the pointing movement would employ overlapping motor representation with the precision grasp programming as the kinematics of the pointing gesture corresponds to—at least

in young children—the kinematics of the opening phase of the precision grasping. On the other hand, the power grasping has more similarities, at least at the kinematic level, with the curl actions that, in contrast to the pointing action that is associated with extension of the hand muscles, is more associated with flexion of the hand muscles. Consequently, the congruency effect observed in the current study with forward (extension) and backward (flexion) movements might be partially based on the same underlying neural mechanisms as the corresponding effects associated with the precision and power grip planning reported by Vainio et al. (2013). However, future work should aim to provide more conclusive evidence of this proposal.

Conclusions

While previous research on hand–mouth integration has shown coupling between hand grasping and articulatory gestures, the results reported here demonstrate for the first time that this interplay can also operate at the level of programming horizontal movements of articulatory organs and hand. Hence, the present study shows evidence for the idea that the horizontal hand movements might also—similarly to grasp movements—share common neural resources with mouth movements. As such, the study provides evidence for the mouth-gesture theories. We suggest that people's innate tendency to unintentionally copy their hand gestures by movements of the tongue—as proposed in the mouth-gesture theories—originates from the shared representations in programming of goal states of these two distal effectors with respect to direction planning (the horizontal direction of tongue and manual movement: the present study) and shape planning (the shape of mouth and manual grasp: Vainio et al. 2013).

Acknowledgments We would like to thank Risto Halonen, Laura Altarriba, Laura Haveri, Tuija Tolonen, Mari Mäkelä, and Milla Vestvik for their contribution in carrying out the experiments. The research leading to these results has received funding from the Academy of Finland under Grant Agreement Number 1265610.

References

- Arbib MA (2005) From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav Brain Sci* 28:105–124
- Binkofski P, Buccino G (2004) Motor functions of the Broca's region. *Brain Lang* 89:362–369

- Chieffi S, Secchi S, Gentilucci M (2009) Deictic word and gesture production: their interaction. *Behav Brain Res* 203:200–206
- Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci* 33:269–298
- Fant C (1960) Acoustic theory of speech production. Mouton, The Hague
- Gentilucci M (2003) Grasp observation influences speech production. *Eur J Neurosci* 17:179–184
- Gentilucci M, Campione GC (2011) Do postures of distal effectors affect the control of actions of other distal effectors? Evidence for a system of interactions between hand and mouth. *PLoS ONE* 6(5):e19793
- Gentilucci M, Rizzolatti G (1990) Cortical motor control of arm and hand movements. In: Goodale MA (ed) *Vision and action: the control of grasping*. Ablex, New York, pp 147–162
- Gentilucci M, Benuzzi F, Gangitano M, Grimaldi S (2001) Grasping with hand and mouth: a kinematic study on healthy subjects. *J Neurophysiol* 86:1685–1699
- Glenberg AM, Kaschak MP (2002) Grounding language in action. *Psychon Bull Rev* 9:558–565
- Graziano MSA, Afshar TN (2007) Mapping behavioral repertoire onto the cortex. *Neuron* 56:239–251
- Hewes GW (1973) Primate communication and the gestural origins of language. *Curr Anthropol* 14:5–24
- Higginbotham DR, Isaak MI, Domingue JM (2008) The exaptation of manual dexterity for articulate speech: an electromyogram investigation. *Exp Brain Res* 186:603–609
- Iverson JM, Thelen E (1999) Hand, mouth and brain. *J Conscious Stud* 6:19–40
- Leung E, Rheingold H (1981) The development of pointing as a social gesture. *Dev Psychol* 17:215–236
- Murphy CM, Messer D (1977) Mothers, infants, and pointing: a study of a gesture. In: Schaffer HR (ed) *Studies in mother-infant interaction*. Academic Press, London, pp 325–354
- Paget R (1930) *Human speech*. Harcourt Brace, New York
- Ramachandran VS, Hubbard EM (2001) Synaesthesia—a window into perception, thought and language. *J Conscious Stud* 8:3–34
- Rizzolatti G, Gentilucci M (1988) Motor and visual-motor functions of the premotor cortex. In: Rakic P, Singer W (eds) *Neurobiology of neocortex*. Wiley, New York, pp 269–284
- Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G et al (1988) Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. *Exp Brain Res* 71:491–507
- Sapir E (1929) A study in phonetic symbolism. *J Exp Psychol* 12:239–255
- Serrien DJ, Bogaerts H, Suy E, Swinnen SP (1999) The identification of coordination constraints across planes of motion. *Exp Brain Res* 128:250–255
- Vainio L, Schulman M, Tiippana K, Vainio M (2013) Effect of syllable articulation on precision and power grip performance. *PLoS ONE* 8(1):e53061
- Vainio L, Tiainen M, Tiippana K, Vainio M (2014) Shared processing of planning articulatory gestures and grasping. *Exp Brain Res* 232:2359–2368
- Waters GS, Fouts RS (2002) Sympathetic mouth movements accompanying fine motor movements in chimpanzees (*Pan troglodytes*) with implications toward the evolution of language. *Neurol Res* 24:174–180