Interaction in planning vocalisations and grasping

Mikko Tiainen*[a], Kaisa Tiippana[a], Martti Vainio[b], Naeem Komeilipoor[a], Lari Vainio[a]

Running head: Planning vocalisations and grasping

Affiliations:

[a] Division of Cognitive and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, Siltavuorenpenger 1-5, 00014 University of Helsinki, Finland

[b] Phonetics and Speech Synthesis Research Group, Institute of Behavioural Sciences, University of Helsinki, Siltavuorenpenger 1-5, 00014 University of Helsinki, Finland

* Correspondence to: mikko.o.tiainen@helsinki.fi, +358503523173, Siltavuorenpenger 1A, room 427, 00014 University of Helsinki, Finland

Word count: 8386 (not including title page).

1

Abstract

Previous studies have shown a congruency effect between manual grasping and syllable articulation. For instance, a power grip is associated with syllables whose articulation involves the tongue body and/or large mouth aperture ([kɑ]) whereas a precision grip is associated with articulations that involve the tongue tip and/or small mouth aperture ([ti]). Previously, this effect has been observed in manual reaction times. The primary aim of the current study was to investigate whether this congruency effect also takes place in vocal responses and to investigate involvement of action selection processes in the effect. The congruency effect was found in vocal and manual responses regardless of whether or not the syllable or grip was known a priori, suggesting that the effect operates with minimal or absent action selection processes. In addition, the effect was observed in vocal responses even when the grip was only prepared but not performed, suggesting that merely planning a grip response primes the corresponding articulatory response. These results support the view that articulation and grasping are processed in a partially overlapping network.

Keywords: action planning, grasping, language evolution, speech

Introduction

We use gestures every day in conjunction with speech (McNeill, 1992). Theories on the existence of connections between manual actions and speech have been around for a long time. For example, Darwin (1872) had noted that when people cut with scissors, they tend to clench their jaws as if to mime the scissor movement. According to gestural theories of language evolution, the foundation for spoken language evolved from a gestural communication system (e.g. Arbib, 2005; Hewes, 1973). In this system our ancestors would have communicated their intentions via mimes of actions, objects or other animals. Ramachandran and Hubbard (2001) proposed a somewhat similar idea by suggesting that some articulations are actually mimes of hand actions. For example, words that denote smallness, such as "little" or "teeny", may involve a narrowing of the vocal tract and could be "synkinetic" mimes of a precision grip, which is used to grasp tiny objects.

Chimpanzees have a tendency to perform mouth movements, such as protruding the tongue and lips that mimic hand actions, during fine-motor grasping actions (Waters & Fouts, 2002). Studies of human participants, such as those done by Gentilucci et al. (e.g. Gentilucci, 2003; Gentilucci & Campione, 2011; Gentilucci, Campione, Dalla Volta, & Bernardis, 2009; Gentilucci, Santunione, Roy, & Stefanini, 2004), have shown corresponding interactions in humans: observing and performing manual actions can influence mouth movements and vocalisations. For example, when a person watches a large object being grasped while simultaneously pronouncing a syllable, the mouth aperture is larger than when the grasped object is small (Gentilucci, 2003). Also, when a person articulates an open vowel and grasps an object, the finger aperture is larger than when a closed vowel is articulated (Gentilucci & Campione, 2011).

Recently, our group discovered a novel connection between hand and mouth actions (Vainio, Schulman, Tiippana, & Vainio, 2013). The syllable-grip effect we discovered showed that performing a power grip was faster when it coincided with the pronunciation of the syllable [kɑ] than with the syllable [ti]. When participants pronounced [ti], they performed the precision grip more quickly than when they were pronouncing [kɑ]. We have proposed that these associations demonstrate that certain articulatory

gestures are programmed in a motor network that partially overlaps with grasp motor representations (Vainio, Tiainen, Tiippana, & Vainio, 2014). With consonants, the tongue body is used to block the airflow at the soft palate to produce [k], whereas the tip of the tongue is used at the alveolar ridge to produce [t]. With vowels, the mouth aperture is larger for [ɑ] (an open vowel) than for [i] (a close vowel). We suggested that articulations involving the tongue body and larger mouth aperture be considered as articulatory equivalents of whole hand movements with larger hand apertures, i.e. a power grip. On the other hand, articulations mainly involving the tip of the tongue and smaller mouth apertures could be considered as resembling hand movements that use the tips of the fingers with smaller hand apertures, such as a precision grip. Our interpretation of the effect is based on previous observations associating the aperture of a vocal tract with grasp processes (e.g. Gentilucci & Campione, 2011). However, it would be equally plausible to assume that this effect could be based on the position of the tongue, since in the vowel [i] the tongue is more up and front than in [ɑ] and vice versa. Regardless of whether the effect is associated with the position of the tongue or the openness of articulation, our primary argument remains the same: the effect is based on interactions in articulation movements and hand grasping movements.

The present study: The influence of grip planning on vocalisation

In the present study, two experiments were designed to further investigate the syllable-grip effect. In the original study (Vainio et al., 2013) only manual reactions were recorded, and consequently what is not known is whether specific grasp actions can influence specific vocalisations. Hence, studying the vocal responses was the principal aim of Experiment 1, where we replicated the original paradigm (Vainio et al., 2013) with additional vocalisation recordings. We have previously proposed that the syllable-grip congruency effect reflects an overlap in the planning processes between grasping and articulation. If this turns out to be the case, then it could be assumed that the effect should also be observed in vocal responses.

Besides vocal reaction times, the vocal characteristics of intensity, fundamental frequency (f_0), formant 1 (F1), and formant 2 (F2) were also of interest, as previous research has shown several connections

between hand actions and these vocal characteristics. Gentilucci, Benuzzi, Gangitano, and Grimaldi (2001) have shown that the intensity of vocalisations is higher when larger objects are grasped as opposed to smaller ones. Increase in f_0 has also been associated with large objects, but only in observations of a hand grasping an object (Gentilucci et al., 2009). F1, on the other hand, has been reported to be higher in observations of power grip rather than precision grip grasps (Gentilucci et al., 2009). Finally, when a person brings fruit to the mouth while articulating syllable BA, F2 has been shown to increase when the fruit is an apple rather than a cherry (Gentilucci et al., 2004). It is possible that performing a power grip instead of a precision grip would elicit an effect similar to grasping a large object or observing a hand grasping a large object. Hence, based on these findings, we could expect all these vocal characteristics to increase/be higher when a power grip is performed rather than a precision grip.

Action execution, action selection and the syllable-grip congruency effect

While Experiment 1 investigated whether vocalisation is influenced by simultaneous grip performance, one of the key questions in Experiment 2a was whether the syllable-grip congruency effect can be observed at the level of vocal responses when the grip response is planned prior to the onset of the syllable, yet is not executed. In other words, whether simply preparing, for example, a precision grip would facilitate the vocal response [ti] even though the actual precision grip execution was withheld. If so, then it could be concluded that the effect does not require response execution processes. This question was addressed by adding a manual go/no-go condition to the task. We used upper- and lowercase letters to distinguish between manual go and no-go situations, whereby participants performed the grip only if the syllable was written in either a lower- or an uppercase letter (balanced between participants).

In Experiments 2a and 2b, the other key question was the extent to which the action selection processes are necessary in the syllable-grip congruency effect. It has been assumed for a long time that priming effects in choice reaction time tasks (i.e., a participant has to select between two opposing response

alternatives at the time of priming) are mostly based on cognitive biases that operate within response selection processes (see Hommel, 1996 for a review). In these cases, the effect is observed because the task involves a competition between opposite response alternatives. The prime has a biasing effect on this competition so that the response option that is compatible with the prime receives competitive advantage, resulting in facilitation of the compatible response alternative and inhibition of the incompatible response alternative. This view is also in line with the affordance competition hypothesis (Cisek, 2007) according to which several responses can be simultaneously competing for which one will be executed. The processes that select the most suitable one for a given situation is largely affected by biasing influences from other simultaneously ongoing sensory and/or motor processes such as processing affordances of the current environment (Cisek & Kalaska, 2010).

In this vein of thought, the priming effects that are observed in simple reaction time tasks (i.e., the required response is known in advance) are commonly attributed to anatomical factors. In these cases, the effect does not rely on cognitively based response selection biases but is rather based on anatomical overlap between representations that are processing the prime and the response. As an example of response priming effect observed in a simple reaction time task, Brass, Bekkering, and Prinz (2001) found that pre-defined finger movements were facilitated when the participants watched a video of a finger movement and the finger in the stimuli was congruent with the response finger. It was suggested that this reaction time effect was observed because the same motor representations that are involved in planning these finger movements were also involved in processing the perceived finger movements leading to the priming effect.

The effect that perhaps has most commonalities with the syllable-grip congruency effect is the size-grip congruency effect originally reported by Ellis and Tucker (2000; Tucker & Ellis, 2001). Similar to the syllable-grip congruency task, in the size-grip congruency task, participants are required to select the precision or power grip for the response according to some stimulus property such as a colour (Ellis, Tucker, Symes, & Vainio, 2007). The responses are made faster if the size of the stimulus is compatible

6

with the grip type required for the response (e.g., small object – precision grip). Importantly for the objectives of the current study, this size-grip congruency effect is entirely eliminated if the response grip is cued 500 ms before the object onset (Tucker & Ellis, 2001; Experiment 2). According to the above mentioned view, this suggests that the size-grip congruency effect is not based on overlap between the representations that are involved in planning grasp responses and processing size information of the stimulus. Rather the effect is more likely based on quite abstract stimulus-response mappings that bias the response selection process.

We have previously suggested that the syllable-grip congruency effect is based on partial overlap between networks that are involved in planning articulatory gestures and manual grasps (Vainio et al., 2013). However, this suggestion would be invalidated if the syllable-grip congruency effect is eliminated when the pressure for response selection is removed from the task – similarly to the size-grip congruency effect, as explained above. Experiment 2 will investigate this issue by introducing a pre-cue for the grip response.

Experiment 1

In Experiment 1, participants were asked to pronounce a syllable ([kɑ] or [ti]) presented on a computer screen whenever its colour changed from grey to green or blue and simultaneously perform a precision grip or a power grip, the grip being selected on the basis of the colour. In comparison with the original study (Vainio et al., 2013) in which only manual responses were recorded, vocal responses were recorded in this experiment as well in order to investigate whether the syllable-grip congruency effect can also be observed in vocal responses. In addition to investigating the influence of syllable-grip congruence on vocal response times, we also investigated whether the vocal characteristics of intensity, f_0, F1, and/or F2 could be influenced by grip performance.

Methods

Participants

Seventeen Finnish-speaking, right-handed (according to a self-report) volunteers (4 male) were recruited for the study. The participants were between 20 and 40 years of age (the mean age being 24.1 years). All reported normal or corrected-to-normal vision and hand-motor functioning and no language disorders. Written, informed consent was acquired from all participants, and all were offered a movie ticket or course credit as a reward for participation. The study was approved by the Ethical Review Board in the Humanities and Social and Behavioural Sciences at the University of Helsinki.

Equipment

The experiment was conducted in a dimly-lit sound-attenuated room. The participants sat in front of a 21" LCD computer screen wearing a head-mounted microphone and holding two response devices in their right hand. The devices were marked with blue and green tape. Both were equipped with an inlaid micro-switch. The cube-shaped precision grip device measured 1×1×0.7 cm and the cylinder-shaped power grip device was 12 cm in length and 3 cm in diameter. As the switches were depressed on each device, there was noticeable tactile feedback. Vocal recording levels were calibrated individually for each participant at the beginning of the experiment. Stimulus presentation and sound recording were done with Presentation® software (Version 16.1, www.neurobs.com).

Stimuli and procedure

The syllables used in the study were [kɑ] and [ti], pronounced short and presented visually according to Finnish orthography (KA and TI). They were written in capital letters with KaiTi font. The font was chosen as it is a mono-spaced sans font. Both syllables were approximately 1.5 degrees of visual angle in height, KA was 2.3 deg in width and TI 1.1 deg in width at a viewing distance of 75 cm. The trial structure was such that a blank screen was displayed for 2000 ms at the beginning of each trial. Then the syllable was shown in the centre of the screen for 400 ms in a light grey colour. Next the syllable

changed to either blue or green, which acted as the participant's go signal (Fig. 1). The participants were instructed to respond to the green colour using the device marked with green tape and to the blue colour with the device marked with blue tape; the response was to be made as quickly as possible. Eight participants responded to green syllables with the precision grip and nine responded to blue syllables with the precision grip. The syllable remained on view for 2000 ms or until a response was made. In addition, the participants were instructed to pronounce the presented syllable simultaneously with the grip response. Erroneous manual responses were followed by a short "beep" tone. Each syllable was presented 30 times in each colour, resulting in 120 trials in total. The experiment began with a training session, and all participants were given enough time to practise until they felt comfortable with the task and the experimenter was sure that the participant could perform the task fluently (approximately 1-2 minutes). The experimenter was able to follow the experiment on a separate screen outside the recording studio and was listening to the vocal responses at all times.

< Figure 1 about here >

Data and statistical analysis

Vocal data were analysed using Praat (v. 5.3.49, www.praat.org). Onsets and offsets were first located individually for each trial as the first observable peak in the acoustic signal for the consonant burst. 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 of the third quarter of the vocalisation area. All reaction time data were cut for each participant so that values two standard deviations smaller or larger than the mean were cut off. The data were subjected to a two-way repeated measure ANOVA with factors of the syllable (i.e. [kɑ] and [ti]) and the grip (precision and power) for each measure of interest. These measures were manual reaction times, vocal reaction times, speech intensity, f_0, F1, and F2. Error rates were analysed in a manner similar to that for the reaction time data. Post-hoc pairwise tests were performed using a Bonferroni correction for multiple comparisons. The effect size was estimated using a partial-eta-squared statistic.

9

In order to confirm the findings with another reaction time cut off, we did a secondary reaction time trimming where reaction times of each subject for different conditions were first separated and then values two standard deviations smaller or larger than the mean for the particular condition were cut off. Results of the statistical analysis for the secondary reaction time trimming are reported in the footnotes for only the most interesting grip-syllable interactions. They confirmed the findings of the primary analysis.

Results

With the manual reaction times, there was a significant interaction between grip and syllable [$F(1,16)=16.73$, $p=.001$, $\eta_p^2=0.51$][a]. Pairwise analysis showed that the power grip was performed faster when [kɑ] was pronounced than when [ti] was pronounced (411 vs 437 ms, respectively, $t(16)=4.89$, $p=.013$), and the precision grip was performed faster when [ti] was pronounced than when [kɑ] was pronounced (403 vs 437 ms, respectively, $t(16)=2.79$, $p<.001$). Importantly, with the vocal reaction times a similar significant interaction between syllable and grip was found [$F(1,16)=26.48$, $p<.001$, $\eta_p^2=0.62$][a]. The syllable [kɑ] was pronounced faster when a power grip was performed than when a precision grip was performed (477 vs 512 ms, $t(16)=3.70$, $p=.002$), and [ti] was pronounced faster when a precision grip was performed (487 vs 521 ms, $t(16)=3.72$, $p=.002$). There were no other significant effects for vocal or manual reaction times. The results are summarised in Figure 2.

For the acoustic properties of intensity, f_0, F1, and F2, main effects of the syllable were found for each property: [kɑ] was pronounced louder than [ti] [$F(1,16)=6.98$, $p=.018$, $\eta_p^2=0.30$, 79.9 vs 78.8 dB,], f_0 was higher for [ti] than for [kɑ] [183 vs 178 Hz, $F(1,16)=44.48$, $p<.001$, $\eta_p^2=0.74$], F1 was higher for [kɑ] than for [ti] [639 vs 354 Hz, $F(1,16)=472.52$, $p<.001$, $\eta_p^2=0.97$], and F2 was higher for [ti] than for [kɑ] [2566 vs 1117 Hz, $F(1,16)=364.26$, $p<.001$, $\eta_p^2=0.96$]. There was also a main effect of grip on intensity [$F(1,16)=9.36$, $p=.007$, $\eta_p^2=0.37$] as well as an interaction between syllable and grip

$[F(1,16)=4.65, p=.047, \eta_p^2=0.23]$. The intensity was louder when a power grip was performed (79.5 dB) than when a precision grip was performed (79.3 dB). The interaction showed that when [ti] was pronounced, the difference between grips was significant ($t(16)=5.33$, $p<.001$), with the power grip associated with louder vocalisations than the precision grip (79.0 vs 78.6 dB). When [kɑ] was pronounced, the difference between grips was not significant (power 80.0 vs precision 79.9 dB, $t(16)=0.60$, $p=.554$). The interaction thus showed that the main effect arose largely from a difference in grips when [ti] was pronounced. There were no other significant predictors of phonetic properties.

< Figure 2 about here >

Overall, the error rate was small, 2.3% for the manual responses and 0.2% for the vocal responses. Owing to the extremely low error rate on vocal responses, these were not analysed. Only three participants made any vocal errors. Because the manual error data were skewed, a Wilcoxon Signed-Ranks Test was used to compare different conditions. These results partly reflected those of the reaction times. When the required grip was a precision grip, more errors were made when the syllable was [kɑ] than when it was [ti] (mean error rates 3.1% vs 1.7%, $Z=-1.96$ $p=.051$). When the required grip was a power grip, the difference was not significant (mean error rates 3.5% vs 1.4%, $Z=-1.49$ $p=.138$).

Discussion

Experiment 1 showed that pronouncing [kɑ] was associated with faster power grip responses and [ti] with faster precision grip responses, replicating the findings of Vainio et al. (2013). It further showed that the correspondence also applies to vocal responses; [kɑ] was pronounced more quickly when executing a power grip and [ti] was pronounced more quickly when executing a precision grip. The effect with vocal responses supports the view that the syllable-grip congruency effect reflects an overlap in the planning processes between grasping and articulation.

Experiment 2

Experiment 2 consisted of two blocks where we explored to what extent the syllable-grip effect relies on action selection processes, by slightly modifying the original paradigm (Vainio et al., 2013). In Experiment 2a, we provided a cue for the required grip type. It preceded the onset of the syllable target to which the articulatory and manual response had to be performed. If the effect is observed even when the task has minimal action planning requirements (i.e. the response is known a priori), it can be assumed that the interconnectedness between the two effectors should be exceptionally tight, as explained in the Introduction. Therefore, if the syllable-grip effect were observed in this pre-cue condition, it would support the view that the effect is based on partial overlap between networks that are involved in planning articulatory gestures and manual grasps.

We added a manual go/no-go condition to the design in order to minimise response anticipations, i.e. to avoid situations in which manual responses are performed at the onset of the syllable presentation without actually processing the syllable. We used upper- and lowercase letters as the go/no-go signal. That is, vocal responses were always performed, but manual responses were performed only when the syllable was written, for example, in capital letters. Using this design, we were also able to investigate whether the syllable-grip congruency effect can be observed in vocal responses even when the manual grip is planned but not executed. Indeed, one of the primary objectives of Experiment 2a was to investigate whether the congruency effect can be observed in vocal responses, similarly to Experiment 1, even when the manual grip response is planned but not executed. If the congruency effect were observed even in this condition, it could be concluded that action execution is not necessary condition for observing the syllable-grip effect.

Experiment 2b was performed in the same session as Experiment 2a. Because the primary objective of Experiment 2 was to determine whether providing a pre-cue for the manual response would influence the syllable-grip congruency effect, Experiment 2b was carried out to provide a comparison to the results of Experiment 2a. In contrast to Experiment 2a in which a pre-cue was provided for the manual

responses, in Experiment 2b no cues were presented, and the syllable was simply displayed onscreen in green or blue to signal the appropriate grip. We kept the manual go/no-go condition in this experiment as well, so that the comparison with Experiment 2a would be as straightforward as possible.

Methods

Participants

Twenty-three Finnish-speaking volunteers (5 male; one participant was left-handed according to self-report) participated in Experiment 2. The participants were between 18 and 29 years of age (the mean age was 24.6 years). All reported normal or corrected to normal vision, normal hand motor functioning and no known language disorders. Written informed consent was acquired from all participants, and all were offered a movie ticket as a reward for partaking. The study was approved by the Ethical Review Board in the Humanities and Social and Behavioural Sciences at the University of Helsinki.

Equipment, stimuli and procedure

The equipment was the same as in Experiment 1. The syllables were also the same in all experiments. In Experiment 2a, a blank screen was first displayed for 2000 ms at the beginning of each trial. A colour cue followed the blank screen. The cue was either a green or a blue circle. Colour pairing was balanced so that 13 participants responded to green stimuli with a precision grip and 10 participants reacted to blue stimuli with the precision grip. The cue was displayed for 400 ms, followed by a 200 ms blank screen and then the syllable [kɑ] or [ti] written in black, either in upper- or lowercase letters (see Fig. 1 for an overview). The uppercase orthographic syllables were approximately 1.5 degrees of visual angle in height, and KA was 2.3 deg in width and TI 1.1 deg in width at a viewing distance of 75 cm. The lowercase syllables were 1.1 degrees of visual angle in height and approximately the same width as their uppercase counterparts. The upper- and lowercase conditions were included as manual go/no-go stimuli. The response mapping was such that 13 participants responded with the grip devices only to syllables written in uppercase letters and 10 participants responded only to syllables written in lowercase letters.

In the go trials, participants responded as quickly as possible with the grip device matching the colour of the cue when the syllable was presented on screen while simultaneously pronouncing it. In the no-go trials they only pronounced the syllable when it was presented, withholding any manual response. The stimuli remained in view for 2000 ms or until a manual response was made.

The setup and procedure were the same in Experiment 2b as in 2a, with the following differences: instead of the initial colour cue, a black fixation cross was presented at the beginning, and the cue for the manual grip appeared simultaneously with presentation of the syllable as the colour of the text, which was either blue or green (Fig. 1). The response mapping was the same as in Experiment 2a. The order of Experiments 2a and 2b was counterbalanced among the participants. Each stimulus combination was presented 30 times, resulting in 240 trials (30 × 2 grips × 2 colours × 2 letter sizes) in each experiment block. With the two blocks, the whole experiment session lasted around 60 minutes including preparation. Feedback, practice and monitoring were the same as in Experiment 1.

Data and statistical analysis

The data were analysed as in Experiment 1 with one exception – the between-subjects factor of response mapping (i.e. the go stimulus for manual response was either upper- or lowercase letters) was added to the ANOVA, making it a mixed-design. We also compared the manual and vocal reaction time effects from Experiment 2b to those of 2a by first subtracting the incongruent reaction time means from the congruent for each syllable (e.g. precision [ti] – power [kɑ]) and then combining these to form one effect size variable for both go and no-go conditions. These were then analysed with a paired samples t-test. Go and no-go data were analysed separately to keep the analyses as simple as possible and also comparable to Experiment 1. Finally, the results of the statistical analysis for the syllable-grip interactions with a secondary reaction time trimming are reported in the footnotes, as in Experiment 1.

Results of Experiment 2a

"Go" conditions

In the manual reaction times there was significant interaction between syllable and grip $F(1,21)=16.64$, $p=.001$, $\eta_P^2=0.44$[b]. Pairwise analysis showed that the precision grip was faster when the syllable was [ti] than when it was [kɑ] (463 vs 532 ms, respectively, $t(22)=4.37$, $p<.001$), and the power grip was faster when the syllable was [kɑ] rather than [ti] (473 vs 524 ms, $t(22)=3.16$, $p=.005$). The interaction between syllable and response mapping was also significant [$F(1,21)=10.59$, $p=.004$, $\eta_P^2=0.34$]. When participants were reacting to lowercase letters, they responded faster to the syllable [ti] than to [kɑ] (468 vs 498 ms, $t(22)=3.54$, $p=.002$). When participants were reacting to capital letters, the difference between the syllables was not significant ([kɑ] 506 vs [ti] 513 ms, $t(22)=0.90$, $p=.380$). Similarly, the interaction between grip and response mapping was significant [$F(1,21)=6.5$, $p=.019$, $\eta_P^2=0.24$]. Reactions with a precision grip were almost significantly faster than those with a power grip in reactions to lowercase letters (475 vs 490 ms, $t(22)=2.00$, $p=.058$). As for reactions to capital letters, reaction times between grips did not differ significantly (power 505 vs precision 515 ms, $t(22)=1.61$, $p=.123$). There was a marginally significant main effect of the syllable: $F(1,21)=4.25$, $p=.052$, $\eta_P^2=0.17$; reactions were faster in general when the syllable was [ti].

In the vocal reaction times, significant interaction between syllable and grip was found, $F(1,21)=14.79$, $p=.001$, $\eta_P^2=0.41$[b]; [kɑ] was pronounced faster when a power grip was performed than when a precision grip was performed (530 vs 587 ms, $t(22)=3.75$, $p=.001$), and [ti] was pronounced faster when a precision grip was performed than when a power grip was performed (530 vs 578 ms, $t(22)=3.69$, $p=.001$). Similar to the result of manual reaction times, there was also significant interaction between syllable and response mapping [$F(1,21)=7.87$, $p=.011$, $\eta_P^2=0.27$]. As for responses to lowercase letters, the reaction times were faster when [ti] was pronounced than when [kɑ] was pronounced (530 vs 555 ms, $t(22)=2.57$, $p=.018$); in responses to uppercase letters the difference was not significant ([kɑ] 561 vs [ti] 572 ms, $t(22)=1.33$, $p=.199$). The interaction between grip and response mapping was significant [$F(1,21)=11.42$, $p=.003$, $\eta_P^2=0.35$]. In responses to uppercase letters while simultaneously performing a power grip, the reaction times were faster than when a precision grip was performed (559 vs 574 ms,

t(22)=3.29, *p*=.003). In responses to lowercase letters, reaction times were not significantly different between grips (power 547 vs precision 538 ms, *t*(22)=1.61, *p*=.123). The results for Experiment 2a are summarised in Figure 3. The vocal characteristic results for both go and no-go conditions are reported in the online Supplemental Material Table 1. These revealed no grip related effects, except for a significant three-way interaction (syllable × grip × response mapping, p=.046) on intensity. However, none of the pairwise comparisons were significant

< Figure 3 about here >

Overall, the error rate was small in the go trials: 3.0% for the manual responses and 0.6% for the vocal responses. Owing to the very low vocal error rates, only manual errors in go trials were analysed. Because the manual error data was skewed, a Wilcoxon Signed-Ranks Test was used to compare the effect of syllable in different grip conditions. When the required grip was a precision grip, more errors were made when the syllable was [kɑ] than when it was [ti] (mean error rates 4.0 vs 0.8%, *Z*=-2.80, *p*=.005); when the required grip was a power grip, more errors were made if the paired syllable was [ti] than when it was [kɑ] (mean error rates 5.4 vs 1.7%, *Z*=-2.61, *p*=.009).

"No-go" conditions

This analysis of vocal responses revealed an interaction of syllable and grip similar to the go trials, $F(1,21)=8.29$, $p=.009$, $\eta_p^2=0.28^b$. As in the go trials, when participants were cued to react with a power grip, they pronounced [kɑ] faster than when they were cued to react with a precision grip (510 vs 525 ms, *t*(22)=2.86, *p*=.009), whereas [ti] was pronounced faster when the participant was cued to perform a precision grip rather than a power grip, although this difference only approached significance (511 vs 518 ms, *t*(22)=1.81, *p*=.084). It should be remembered that, although there was always a cue for one of the grips, no manual responses were actually made in these trials. In the no-go trials the error rate for vocal responses was only 0.04% and thus these errors were not analysed.

Results for Experiment 2b

"Go" conditions

With manual reaction times, we again replicated the syllable-grip interaction effect, $F(1,21)=63.27$, $p<.001^c$, with a noticeably large effect size $\eta_p^2=0.75$. The precision grip was performed faster when [ti] was pronounced than when [ka] was pronounced (595 vs 652 ms, $t(22)=7.68$, $p<.001$), and the power grip was performed faster when [ka] was pronounced than when [ti] was pronounced (604 vs 664 ms, $t(22)=5.24$, $p<.001$). A comparison of the manual syllable-grip interaction effect sizes in Experiments 2a and 2b showed no significant difference between the two, $t(22)=0.11$, $p=.914$ (2a: 59 ms vs 2b: 58 ms). The interaction between syllable and response mapping was significant, $F(1,21)=5.89$, $p=.024$, $\eta_p^2=0.22$. In responses to uppercase letters, reaction times were faster when the syllable was [ka] than when it was [ti] (618 vs 631 ms, $t(22)=1.71$, $p=.101$), whereas reaction times to lowercase letters were faster when the syllable was [ti] rather than [ka] (627 vs 642 ms, $t(22)=1.72$, $p=.100$). Although neither of these syllable differences was significant, their reverse direction was the cause of the significant interaction effect. Unlike in Experiment 2a, the interaction between grip and response mapping was only approaching significance, $F(1,21)=3.76$, $p=.066$, $\eta_p^2=0.15$. The main effect of grip was significant, $F(1,21)=4.73$, $p=.041$, $\eta_p^2=0.18$. Precision grip reaction times were faster than power grip times (624 vs 635 ms).

Vocal reaction time results showed a highly significant interaction of syllable and grip, again with an exceptionally large effect size, $F(1,21)=49.95$, $p<.001$, $\eta_p^2=0.70^c$: [ka] was pronounced faster when a power grip was performed than when a precision grip was performed (637 vs 677 ms, $t(22)=4.59$, $p<.001$), and [ti] was pronounced faster when a precision grip was performed than when a power grip was performed (638 vs 686 ms, $t(22)=6.78$, $p<.001$). A comparison of the vocal syllable-grip interaction effect size in Experiments 2a and 2b showed no difference between the two, $t(22)=0.60$, $p=.556$ (2a: 52 ms vs 2b: 44 ms). The interaction of syllable and response mapping was significant, $F(1,21)=11.36$, $p=.003$, $\eta_p^2=0.35$. In responses to uppercase letters, the difference between syllables was significant;

[ka] was pronounced more quickly than [ti] (644 vs 664, $t(22)=3.42$, $p=.003$). In responses to lowercase letters, the difference was not significant ([ti] 659 vs [ka] 669 ms, $t(22)=1.48$, $p=.154$). Unlike in Experiment 2a, the interaction between grip and response mapping was only marginally significant, $F(1,21)=4.24$, $p=.052$, $\eta_p^2=0.17$. The results for Experiment 2a are summarised in Figure 4. The vocal characteristic results for both go and no-go conditions are reported in the online Supplemental Material Table 1. These revealed no significant grip effects.

< Figure 4 about here >

The error rate was 3.9% in the go trials for the manual responses and 0.5% for the vocal responses. Hence, only manual errors of the go trials were analysed. Because the manual error data was skewed, a Wilcoxon Signed-Ranks Test was used to compare the effect of syllable in different grip conditions. When the required grip was a precision grip, more errors were made when the syllable was [ka] than when it was [ti] (mean error rates 5.2 vs 1.4%, $Z=-3.33$, $p=.001$); when the required grip was a power grip, more errors were made if the paired syllable was [ti] compared to [ka] (mean error rates 8.2 vs 1.6%, $Z=3.77$, $p<.001$).

"No-go" conditions

The interaction of syllable and grip was marginally significant in vocal reaction times, $F(1,21)=3.97$, $p=.059$, $\eta_p^2=0.16^c$. Interestingly, the interaction effect seemed reversed here, compared to previous vocal reaction time results. When the colouring of the letters was associated with a precision grip, [ka] was pronounced faster than when the colouring was associated with a power grip, although not significantly (555 vs 561 ms, $t(22)=0.93$, $p=.364$). On the other hand, when the syllable's colouring was related to a power grip, [ti] was pronounced more quickly than when the colouring related to a precision grip, but again, not significantly (556 vs 564 ms, $t(22)=1.61$, $p=.121$)[d]. In the no-go trials the error rate for vocal responses was only 0.1% and thus these errors were not analysed.

Discussion

The results of Experiment 2a demonstrated that the syllable-grip congruency effect can be observed in vocal responses even when the grip was planned beforehand but not actually executed, suggesting that action execution is not required condition for the effect. In addition, the results of Experiment 2a showed that when a manual action was planned beforehand (i.e., the manual response was pre-cued), the congruency effect was observed, reflecting a biasing influence of vocal processes on manual processes. This indicates that the syllable-grip congruency effect is preserved also in a situation where manual action selection is not required after onset of the priming stimulus (i.e., the syllable). Moreover, the results of Experiment 2b showed that the syllable-grip congruency effect is similar regardless of whether the grip type is pre-cued (Experiment 2a) or not (Experiment 2b). The evidence support the view that the overlapping motor representations might be indeed involved in planning action of both effectors.

General discussion

Previous research has shown that precision and power grip responses are made faster when the participant pronounces a syllable that is congruent with the required grip type (e.g. [ti] – precision grip) simultaneously with the grip performance (Vainio et al., 2013). In addition to replicating this effect in manual reaction times, this syllable-grip congruency effect was similarly observed in all experiments for vocal reaction times. The syllable [kɑ] was associated with faster power grip responses and was also pronounced faster when paired with a power grip. In contrast, [ti] was associated with faster precision grip responses and was pronounced faster when coupled with a precision grip. These results support the bi-directionality between mouth and hand actions, similar to what the studies by Gentilucci et al. (2001; Gentilucci & Campione, 2011) have shown, and extends them by showing the specificity of these connections: a specific grip type can be associated with a specific articulatory gesture.

One of the key findings of the current study was that the syllable-grip congruency effect was present for vocal reaction times even when a manual grip was prepared, yet not actually executed. In other words,

the effect was found in manual no-go trials in Experiment 2a in which the participants were required to withhold the prepared manual response according to letter case information. Since the participants were aiming to synchronize the two responses in order to follow the task instructions, the congruency effect observed in vocal reaction times of Experiment 1 might have been just a reflection of the manual congruency effect. The fact that the effect was observed in vocal responses even when grip was prepared but not executed shows that the syllable-grip congruency effect in vocal reactions is not, however, just a by-product of the syllable-grip congruency effect related to manual responses, but rather a genuine motor priming effect triggered by motor planning processes of hand grip.

As mentioned in the Introduction, if the response priming effect is observed even when the required response is known before the onset of the stimulus on which the response is executed, it is likely to be based on tight interconnectedness, or even partial overlap between the representations linked to the prime and response. The syllable-grip congruency effect was present in vocal reaction times even when the vocal response was known beforehand (Experiment 1) and in manual reaction times even when the manual response was known beforehand (Experiment 2a). Consequently, we propose that the effect does not reflect some learned abstract associations between articulatory and grip representations, but rather is more likely to reflect functionally relevant, partially shared mechanisms that operate for planning articulatory gestures and grasp actions (e.g., planning the goal shapes of the two distal effectors). As such, our interpretation of the effect can be assumed to support the gestural theories of language evolution (e.g. Arbib, 2005; Hewes, 1973). Furthermore, it has been previously suggested that representations of certain articulatory gestures and manual grasp actions can be tightly integrated at the functional and anatomical levels (Gentilucci & Corballis, 2006; Vainio et al., 2014). For example, the same premotor neurons program grasp actions performed with hand and mouth (Rizzolatti et al., 1988). Although it is difficult to address the actual neural substrate for the syllable-grip congruency effect using behavioural methods, the present study supports the view – at least in the light of the above proposed theoretical stance - that the effect is based on partial overlap between networks that are involved in planning articulatory gestures and manual grasps (Vainio et al., 2013).

20

In general, vocal characteristics were not influenced by manual performance in the study. This was somewhat unexpected because Gentilucci and colleagues (2001; 2009; 2004), for example, have reported several vocal modulations during manual grasping or grasp observation. One explanation for this could be that here, the movement required to perform the manual grips was very small basically reflecting just a terminal phase of grasp actions. In contrast, the grasps in the studies by Gentilucci et al. (e.g. 2004) have usually been large reach-to-grasp types of movements. It could be that in order to see modulations in the vocal properties, larger and more sustained hand movements are needed. We did find an effect of intensity modulated by grip, but only in Experiment 1, where power grip was associated with louder vocalisations when the participants pronounced the syllable [ti]. Gentilucci et al. (2009; 2004) have also reported grips affecting vocalisations, where increasing object size has been associated with increased intensity of vocalisations when participants viewed objects being grasped. We could speculate that performing a power grip also requires the speaker to expend more effort with simultaneous articulations in incompatible situations, where the grip does not match the articulation.

Additionally, we found a secondary effect in relation to letter case. Reactions were relatively rapid with the precision grip when the go-stimulus was a syllable written in lowercase letters and with the power grip when it was written in uppercase letters. However, this effect was significant only in Experiment 2a, although a similar trend was observed in Experiment 2b. We would consider this to be a relatively trivial finding, since the subjects were specifically asked to make a distinction between the letter cases and, in Finnish, uppercase is referred to, literally, as "big" letters and lowercase as "small" letters. The participants were thus in principle asked to respond manually according to the big or small categories of the stimuli. This finding can be assumed to be in line with previous studies that have shown a power grip to be associated with large objects and a precision grip with small objects (Tucker & Ellis, 2001). There was also another letter case-related effect: an interaction between letter case and syllable. The syllable [kɑ] was pronounced relatively rapidly with uppercase letters and the syllable [ti] with lowercase letters. We assume that this effect reflects the same processes as the letter case effect related

to grip responses, as our main findings already suggest a tight connection between grips and articulations. That is, for example, when responding to "small" stimuli (i.e., lowercase letters), not only the grip responses congruent with this stimulus type (i.e., the precision grip) are produced relatively rapidly but also articulations congruent with this stimulus type (i.e., [ti]) are produced relatively rapidly.

Finally, it has been recently found that a spatial manual task affects a spatial working memory task performance but not a verbal working memory task performance (Spiegel, Koester, & Schack, 2013). In a future study, it could be interesting to utilize the findings of the current study in a working memory task to see if performing these syllable related grip actions could influence performance on a verbal working memory task as well.

In conclusion, the present study reported the syllable-grip congruency in vocal responses for the first time. The study also found the congruency effect to be sustained in situations such as when the action selection stage was eliminated and also when the action had only been selected, but was not actually performed. This is evidence that the effect can arise solely from the action planning, yet it can also be observed when action selection requirements of the task are minimal. These findings are a showcase of the robustness of the syllable-grip congruency effect and are in line with our previous suggestions that manual and speech motor mechanisms share a partly overlapping network.

Acknowledgements

References

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences, 28*(02), 105-124.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta psychologica, 106*(1), 3-22.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1485), 1585-1599.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience, 33*, 269-298.

Darwin, C. (1872). *The expression of emotions in animals and man*. London: Murray.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British journal of psychology, 91*(4), 451-471.

Ellis, R., Tucker, M., Symes, E., & Vainio, L. (2007). Does selecting one visual object from several require inhibition of the actions associated with nonselected objects? *Journal of Experimental Psychology: Human Perception and Performance, 33*(3), 670.

Gentilucci, M. (2003). Grasp observation influences speech production. *European Journal of Neuroscience, 17*(1), 179-184.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology, 86*(4), 1685-1699.

Gentilucci, M., & Campione, G. C. (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., Campione, G. C., Dalla Volta, R., & Bernardis, P. (2009). The observation of manual grasp actions affects the control of speech: A combined behavioral and Transcranial Magnetic Stimulation study. *Neuropsychologia, 47*(14), 3190-3202.

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: a gradual transition. *Neuroscience & Biobehavioral Reviews, 30*(7), 949-960.

Gentilucci, M., Santunione, P., Roy, A. C., & Stefanini, S. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *European Journal of Neuroscience, 19*(1), 190-202.

Hewes, G. W. (1973). Primate Communication and the Gestural Origin of Language. *Current Anthropology*, 5-24.

Hommel, B. (1996). SR compatibility effects without response uncertainty. *The Quarterly Journal of Experimental Psychology: Section A, 49*(3), 546-571.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*: University of Chicago Press.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of consciousness studies, 8*(12), 3-34.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental brain research, 71*(3), 491-507.

Spiegel, M. A., Koester, D., & Schack, T. (2013). The functional role of working memory in the (re-) planning and execution of grasping movements. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1326.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual cognition, 8*(6), 769-800.

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. *Experimental brain research, 232*(7), 2359-2368.

Waters, G. S., & Fouts, R. S. (2002). Sympathetic mouth movements accompanying fine motor movements in chimpanzees (Pan troglodytes) with implications toward the evolution of language. *Neurological Research, 24*(2), 174-180.

Footnotes

a: Results when reaction times were trimmed by participant and condition: Manual reaction time grip-syllable interaction $F(1,16)=19.45$, $p<.001$, $\eta_p^2=0.55$. Vocal reaction time grip-syllable interaction $F(1,16)=31.90$, $p<.001$, $\eta_p^2=0.67$.

b: Results when reaction times were trimmed by participant and condition: Manual reaction time grip-syllable interaction $F(1,21)=20.64$, $p<.001$, $\eta_p^2=0.50$. Vocal reaction time grip-syllable interaction $F(1,21)=16.58$, $p=.001$, $\eta_p^2=0.44$. No-go vocal reaction time syllable-grip interaction $F(1,21)=8.40$, $p=.009$, $\eta_p^2=0.29$.

c: Results when reaction times were trimmed by participant and condition: Manual reaction time grip-syllable interaction $F(1,21)=75.02$, $p<.001$, $\eta_p^2=0.78$. Vocal reaction time grip-syllable interaction $F(1,21)=71.82$, $p<.001$, $\eta_p^2=0.77$. No-go vocal reaction time syllable-grip interaction $F(1,21)=2.65$, $p=.118$, $\eta_p^2=0.11$.

d: Although the differences were not significant, their reverse directions were the only possible explanation for the interaction. It is possible that the grip response (e.g. the power grip) is initially prepared - to some extent - according to the stimulus colour, even in these no-go conditions. Consequently, the corresponding articulatory representation (i.e. [ka]) is simultaneously partially activated. However, requirement for rapid suppression of the manual activation might lead to simultaneous suppression of the partially activated articulatory representation, resulting in relatively slow vocal responses when the grip and the vocal response are congruent. However, as this interaction was only marginally significant, and not significant with the secondary reaction time trimming, it will not be discussed further in this paper.

Figure captions

*Figure 1.*Trial structures for each of the experiments. In Experiment 1, the syllable was first presented in light grey. When it changed colour, both a manual response (a precision or power grip) and a vocal

response (pronouncing the written syllable) were to be made simultaneously. In Experiment 2a, a coloured circle was first presented, which acted as a cue for the manual response, followed by an empty screen and finally a syllable written in black, at which point both manual and vocal responses were to be made. In Experiment 2b, there were no a priori cues, just a fixation cross followed by an empty screen and then a syllable written in green or blue to indicate the proper grip response. In Experiments 2a and 2b, the letter case of the syllable acted as a go/no-go signal; in no-go trials the manual response was withheld and only a vocal response was made.

Figure 2. Syllable and grip interactions in manual and vocal reaction times in Experiment 1. With manual reaction times, a power grip was performed faster when [kɑ] was pronounced than when [ti] was pronounced; a precision grip was performed faster when [ti] was pronounced than when [kɑ] was pronounced. With vocal reaction times, [kɑ] was pronounced faster when a power grip was performed than when a precision grip was performed, and [ti] was pronounced quicker when a precision grip was performed. * $p<.05$, ** $p<.01$.

Figure 3. Syllable and grip interactions in manual and vocal reaction times in Experiment 2a. In the manual reaction times, even if the grip was known beforehand, the power grip was performed faster when [kɑ] was pronounced than when [ti] was pronounced; the precision grip was performed faster when [ti] was pronounced than when [kɑ] was pronounced. In the vocal reaction times, [kɑ] was pronounced faster when a power grip was performed than when a precision grip was performed; [ti] was pronounced faster when a precision grip was performed. In the no-go trials the vocal reaction time effect was slightly smaller. ** $p<.01$, *** $p<.001$.

Figure 4. Syllable and grip interactions in manual and vocal reaction times in Experiment 2b. In the manual reaction times, the power grip was performed faster when [kɑ] was pronounced than when [ti] was pronounced; the precision grip was performed faster when [ti] was pronounced than when [kɑ] was pronounced. In the vocal reaction times, [kɑ] was pronounced faster when a power grip was performed

than when a precision grip was performed, and [ti] was pronounced faster when a precision grip was performed. In the no-go vocal responses the interaction effect was marginally significant, only in the reverse direction, and none of the pairwise comparisons was significant. ** $p<.01$, *** $p<.001$.