

## Congruency effect between articulation and grasping in native English speakers

Mikko Tiainen<sup>1</sup>, Fatima M. Felisberti<sup>2</sup>, Kaisa Tiippana<sup>1</sup>, Martti Vainio<sup>1</sup>, Juraj Simko<sup>1</sup>,  
Jiri Lukavsky<sup>3</sup>, Lari Vainio<sup>1</sup>

<sup>1</sup> Institute of Behavioural Sciences, University of Helsinki, Finland

<sup>2</sup> Kingston University London, United Kingdom

<sup>3</sup> Institute of Psychology, Czech Academy of Sciences, Czech Republic

{mikko.o.tiainen, kaisa.tiippana, martti.vainio, juraj.simko, lari.vainio}@helsinki.fi,  
f.felisberti@kingston.ac.uk, lukavsky@praha.psu.cas.cz

### Abstract

Previous studies have shown congruency effects between specific speech articulations and manual grasping actions. For example, uttering the syllable [ka] facilitates power grip responses in terms of reaction time and response accuracy. A similar association of the syllable [ti] with precision grip has also been observed. As these congruency effects have been to date shown only for Finnish native speakers, this study explored whether the congruency effects generalize to native speakers of another language. The original experiments were therefore replicated with English participants (N=16). Several previous findings were reproduced, namely the association of syllables [ka] and [ke] with power grip and of [ti] and [te] with precision grip. However, the association of vowels [a] and [i] with power and precision grip, respectively, previously found for Finnish participants, was not significant for English speakers. This difference could be related to ambiguities of English orthography and pronunciation variations. It is possible that for English speakers seeing a certain written vowel activates several different phonological representations associated with that letter. If the congruency effects are based on interactions between specific phonological representations and grasp actions, this ambiguity might lead to weakening of the effects in the manner demonstrated here.

**Index Terms:** speech production, manual actions, grasping

### 1. Introduction

Several studies have shown that speech and manual movements are connected. For example, when a participant is watching a large object being grasped while simultaneously pronouncing a syllable, the mouth is opened more than when watching a smaller object being grasped [1]. Also, when an open vowel is articulated while grasping an object, the finger aperture is wider than when a closed vowel is articulated [2].

Ramachandran and Hubbard have suggested that some articulations are mimes of hand actions [3]. For example, words denoting smallness may involve narrowing of the vocal tract (e.g., “little” or “teeny”). This narrowing could be an analogous movement for the precision grip in which the finger aperture is similarly narrow.

Our group has found specific connections between manual grasp actions and articulations [4]. In this articulation-grip congruency effect performing a power grip was faster when it coincided with the pronunciation of the syllables [ka], [ke] or [ha] than with the syllables [ti], [te] or [hi]. When participants pronounced [ti], [te] and [hi] they performed the precision grip more quickly than when they pronounced [ka], [ke] or [ha].

We proposed that these associations demonstrate that the network for performing articulatory gestures partially overlaps with grasping networks. For [k], the tongue body is used to block the airflow at the soft palate, whereas for [t] the tip of the tongue is used to block airflow at the alveolar ridge. With the vowels [a] and [i], the mouth aperture is larger for [a] (an open vowel) than for [i] (a closed vowel). Like Ramachandran and Hubbard [3] suggested, articulations involving larger mouth aperture could be analogues of whole hand movements with larger finger aperture i.e. a power grip. On the other hand, articulations with smaller mouth apertures could be considered analogues of hand movements with smaller finger apertures, like a precision grip. Likewise articulations involving the tongue body could be similar to hand movements using the palm of the hand, as in power grip, and articulations involving the tip of the tongue could be similar to hand movements that involve the tips of the fingers, as in precision grip.

A connection between meaningless utterances and grasping actions supports the idea that manual and articulatory gestures are connected even at a pre-semantic level; processes involved in planning articulatory gestures and grasp actions are partially overlapped. In addition, in a wider evolutionary context, the articulation-grip congruency effect can be assumed to be in line with those gestural theories of language evolution that assume that spoken language evolved from or along with communication based on manual gestures and that initial milestone for this evolution was evolution of system that enabled imitation of manual manipulation and grasp actions [5, 6].

However, to date, all studies about the articulation-grip congruency effect have been conducted with Finnish participants. It is possible that the reported connections are specific to the Finnish language. Although the syllables themselves are meaningless in Finnish, there is still a possibility that the results somehow reflect semantic or other associations that are specific to the Finnish language.

In this study we addressed the generalizability of the articulation-grip congruency effect to another language, namely English. We used the same paradigm as in Finnish [4] with syllable pairs [ka]-[ti], [ke]-[te] and [a]-[i] to see if, and to what extent, we can observe the same congruency effects in another language. In addition to the original study, vocal responses were recorded as well. We have recently shown that the effect is also present in vocal reaction times [7]. As the connections between speech and hand actions have been proposed to be evolutionally old, we would expect these effects to be universal and not specific to Finnish. Thus, we expected to find an interaction between articulation and grip execution in all blocks; for a consonant change ([ke]-[te]), for a vowel change ([a]-[i]), and when both change ([ka]-[ti]). A failure to replicate the effect in this study would suggest that it reflects semantic associations in a specific language environment.

## 2. Methods

### 2.1. Participants

Sixteen volunteer native English speakers participated in the experiment (3 male). Two participants were from the United States; the rest were British. The mean age was 22.5 years (range 19 to 31 years). All participants were right-handed (one ambidextrous) and reported normal vision. Written informed consent was acquired from all participants who received a £ 6 Amazon voucher as compensation for their time. The study was conducted at Kingston University London, U.K. The study was approved by the Ethical Review Board in Humanities and Social and Behavioural Sciences at the University of Helsinki.

### 2.2. Equipment, stimuli and procedure

The experiment was conducted in a dimly lit sound attenuated room. Participants sat in front of a computer screen wearing a head-mounted microphone and holding the response devices in their right hand (Fig. 1). The response devices were equipped with a micro-switch. The cube-shaped precision grip device measured 1×1×0.7 cm and the cylinder-shaped power grip device was 12 cm long, and 3 cm in diameter. There was noticeable tactile feedback when the switches were pressed. The devices were marked with blue and green tape. Stimulus presentation and sound recording were done with Presentation® software (Version 16.1, www.neurobs.com).

The experiment consisted of three blocks, with one pair of vowels or syllables in each block. Each trial started with a blank screen for 2000 ms, followed by the syllable/vowel written in light grey in the middle of the screen for 400 ms after which it changed color into green or blue (see Fig. 1). The color indicated which grip response was to be performed. Seven participants responded with a precision grip to green stimuli and with a power grip to blue stimuli. For the remaining nine participants this mapping was reversed. The color change acted as the go-signal for both manual and vocal response. The colored stimulus remained on screen until a manual response was made, or maximally for 2000 ms. Erroneous manual responses were followed by a short “beep” tone. The first block consisted of syllables [ka] and [ti], written as ⟨KAH⟩ and ⟨TEE⟩, respectively. The second block consisted of syllables [ke] and [te], written as ⟨KEH⟩ and ⟨TEH⟩. The third block consisted of vowels [a] and [i], written as ⟨AH⟩ and ⟨EE⟩. Each syllable-color combination was presented 30 times, resulting in 360 trials in total (30 × 2 grips × 2 syllables/vowels × 3 blocks). The trials were presented in random order. Participants were

instructed on the pronunciation of the stimuli to ensure similar pronunciation as in previous studies. They were given practice trials before the actual experiment.

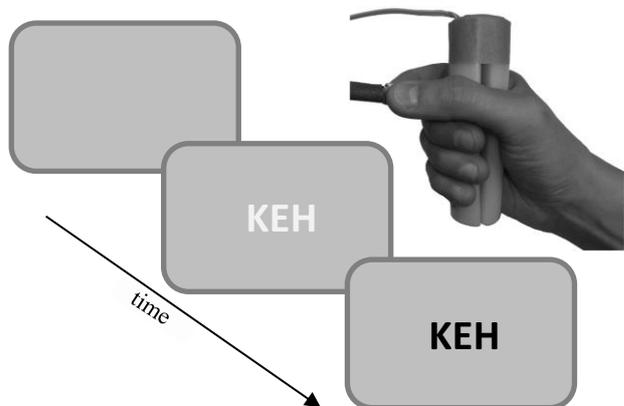


Figure 1: *Left: trial structure of the task. Each trial started with an empty screen, followed by the syllable written in light grey. After 400 ms the syllable changed to either blue or green (black in graph), at which point both the manual and vocal response were to be made. Right: Picture of both grip devices as held by the participants. Note that the precision grip device is held between the thumb and the index finger.*

### 2.3. Data and statistical analysis

Vocal data were analyzed using Praat (5.3.49). Onsets were located individually for each trial as the first observable peak in the acoustic signal for the consonant burst or the vowel onset. Erroneous manual and vocal responses were removed from the data (in total 4.8% of the data). The remaining manual and vocal reaction time data were filtered for each participant so that values two standard deviations smaller or larger than the mean were removed. This left 91.2% of the manual and 90.8% of the vocal data for the final analysis. The data was subjected to separate repeated measures ANOVAs with the factors syllable (e.g. [ka]-[ti]) and grip (precision-power) for manual and vocal reaction times and error rates in each block. Post hoc comparisons were performed by means of t-tests applying a Bonferroni correction. A partial-eta-squared statistic served as effect size estimate. One subjects data for the [ka]-[ti] block had to be rejected due to missing vocal data. Five participants mispronounced the vowel [a] as [æ] on several trials but their data was left in the analysis since [æ] is also an open vowel.

## 3. Results

### 3.1. [ka]-[ti] block

Both manual and vocal reaction time results are presented in Figure 2. For the manual reaction times there was a main effect for grip,  $F(1,14)=15.262, p=.002, \eta^2=0.522$ . Precision grip (631 ms) was performed more quickly than power grip (673 ms). Importantly, the syllable-grip interaction was also significant,  $F(1,14)=13.844, p=.002, \eta^2=0.497$ . When pronouncing [ka], precision grip responses were slower (651 ms) than when pronouncing [ti] (610 ms,  $p=.012$ ). Conversely, when

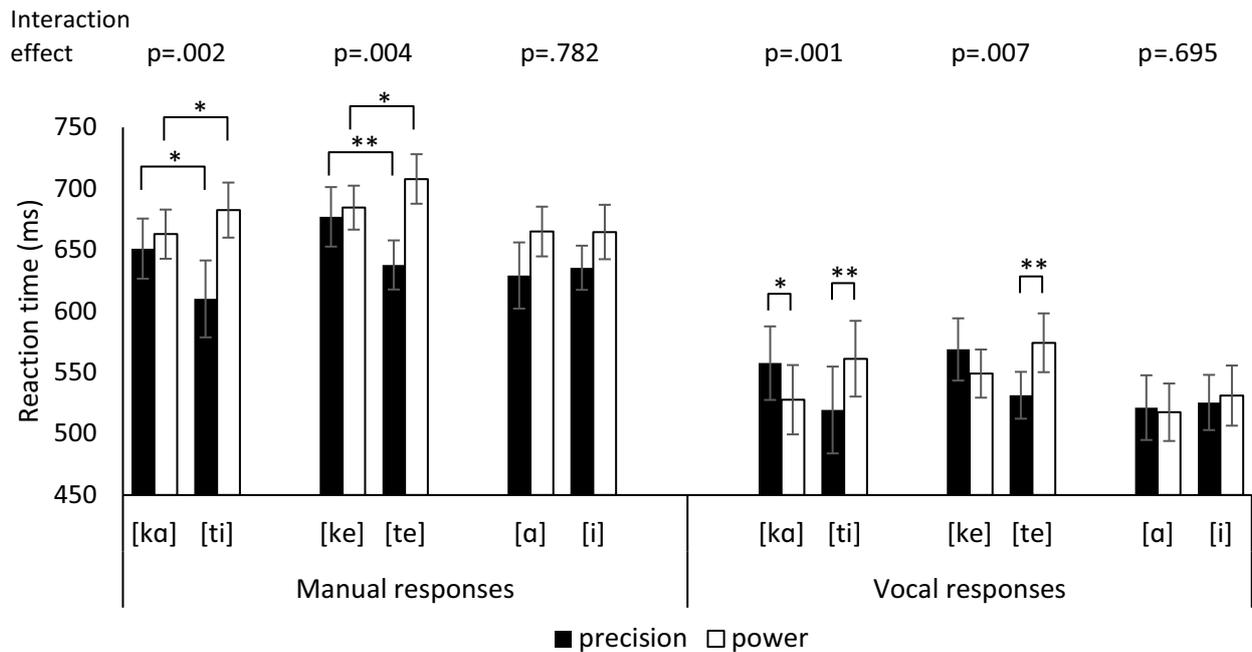


Figure 2: Manual and vocal response results. In manual responses, power grip reaction times when articulating [ka] or [ke] were faster than when articulating [ti] or [te]. Precision grip reaction times were faster when articulating [ti] or [te] than when articulating [ka] or [ke]. In vocal responses, reaction times for [ka] and [te] were faster when simultaneously executing a power grip than when executing a precision grip. Reaction times for [ti] and [te] were faster when performing a power grip than when performing a precision grip. Error bars represent the standard error. \*  $p < .05$ , \*\*  $p < .01$

pronouncing [ka], power grip responses were executed faster (663 ms) than when pronouncing [ti] (682 ms,  $p = .044$ ).

For the vocal reaction times only the syllable-grip interaction was significant,  $F(1,14) = 18.723$ ,  $p = .001$ ,  $\eta^2 = 0.572$ . [ka] was pronounced slower (558 ms) when performing a precision grip than when performing a power grip (528 ms,  $p = .028$ ). Conversely, [ti] was pronounced faster (519 ms) when executing a precision grip than when executing a power grip (561 ms,  $p = .002$ ).

The syllable-grip interaction was also significant when analyzing the error rates,  $F(1,14) = 12.426$ ,  $p = .003$ ,  $\eta^2 = 0.47$ . There were more errors when the required syllable was [ka] and the required manual response was the precision grip (7.8%) than when the syllable was [ti] (1.3%,  $p = .015$ ). In contrast, there were fewer errors when the syllable was [ka] and the manual response was the power grip (3.1%) than when the syllable was [ti] (6.9%,  $p = .011$ ).

### 3.2. [ke]-[te] block

In the [ke]-[te] block, there was a main effect of grip for manual reaction times,  $F(1,15) = 12.939$ ,  $p = .003$ ,  $\eta^2 = 0.463$ . Precision grip responses were faster (657 ms) than power grip responses (696 ms). The syllable-grip interaction was also significant,  $F(1,15) = 11.868$ ,  $p = .004$ ,  $\eta^2 = 0.442$ . Precision grip responses were slower when pronouncing [ke] (677 ms) than when pronouncing [te] (637 ms,  $p = .006$ ). Conversely, power grip responses were faster when pronouncing [ke] (684 ms) than when pronouncing [te] (708 ms,  $p = .016$ ).

For the vocal responses only the syllable-grip interaction was significant,  $F(1,15) = 9.705$ ,  $p = .007$ ,  $\eta^2 = 0.393$ . The difference between grips was not significant when pronouncing

[ke] (precision 569 ms, power 549 ms,  $p = .124$ ). When performing a precision grip, [te] was pronounced faster (531 ms) than when performing a power grip (574 ms,  $p = .006$ ). There were no significant error rate effects.

### 3.3. [a]-[i] block

In the [a]-[i] block, the only significant result was a main effect of grip for manual responses,  $F(1,15) = 13.859$ ,  $p = .002$ ,  $\eta^2 = 0.48$ . Precision grip responses were faster (632 ms) than power grip responses (665 ms). However, the vowel-grip interaction was not significant  $F(1,15) = 0.079$ ,  $p = .782$ ,  $\eta^2 = 0.005$ . There were no significant effects for the vocal responses or error rates.

## 4. Discussion

We found the articulation-grip congruency effect with syllables [ka]-[ti] and [ke]-[te] in native English speakers. In the manual responses, the power grip was executed faster when pronouncing [ka] or [ke] and the precision grip was executed faster when pronouncing [ti] or [te]. In the vocal responses, [ka] and [ke] were pronounced faster when executing the power grip and [ti] and [te] were pronounced faster when the precision grip was executed. In the [ka]-[ti] block the congruency effect was also observed in the error rates, so that more errors were made on the incongruent precision-[ka] and power-[ti] trials than in the congruent power-[ka] and precision-[ti] trials. These results are in line with those observed with Finnish participants [4, 7, 8].

However, the congruency effect was completely absent in the [a]-[i] block. This is in contrast to studies with Finnish participants, for whom the effect is found also with these vowels [4]. So it seems that the congruency effect is not language

specific for consonants, whereas it could be language specific for vowels. In English the correspondence between orthography and pronunciation is much more irregular than in Finnish. It is possible that seeing a certain letter activates several different phonological representations in English speakers, given that seen letters have been proposed to be automatically mapped to corresponding phonetic representations [9]. The current results could reflect that vowel graphemes (e.g., ⟨A⟩ and ⟨E⟩) activate a wider variety of phonological representations than consonant graphemes in English speakers. If the articulation-grip congruency is based on congruencies between specific articulatory gestures and grips as we have previously suggested [4, 7, 8], it may be that the activation of several phonological representations (i.e., articulatory representations, [10]) would cause the congruency effect to vanish. This could mean that the vowel effect is not necessarily specific to Finnish, but its absence is specific to English or any other language with ambiguities between orthography and pronunciation.

However, there are some points regarding the current study that prevent us from making a strong claim about this issue. First of all, the stimuli were visually different from stimuli used in our previous studies, i.e. ⟨AH⟩ and ⟨EE⟩ instead of ⟨A⟩ and ⟨I⟩. This writing form was chosen to facilitate natural pronunciation of the stimuli as [ɑ] and [i], but it could have actually caused some other, yet unknown, interference in the effect. Second, five participants kept mispronouncing one of the vowels, which is a concern since the effect is thought to be based on an overlap of specific motor representations. This raises an interesting possibility; the effect might not be related to the open-close aspect of articulation and grasping, but more to front-back dimension. This is because [æ], which was how the participants mostly mispronounced [ɑ] as, is a front vowel, as is [i]. Indeed, we have recently shown a similar congruency effect between front-back hand movements and articulations [11]. The grip effect could actually be a manifestation of the same phenomenon. Precision grip can be thought of as more of a pointing gesture, where the index finger and thumb are pointing outward. On the other hand, in the power grip all fingers are pointing more backwards, towards the palm of the hand. It has been argued that stimulus-response effects, like the articulation-grip congruency, usually require selection between two opposite alternatives [12]. If the effect indeed operates in the front-back domain instead of the open-close, we would not expect an effect if the responses are both front vowels. The number of participants in the current study was too low to statistically test these speculations. This could be studied more in the future, for example, by pairing a more closed back vowel, such as [o], with [i].

The fact that several participants mispronounced the vowel [ɑ] as [æ] seems to actually support the above proposition that the letters activated other phonological representations, even despite the training on pronunciation received prior to the experiment. Possibly similar phonological activation occurred with the other participants too, but they managed to successfully inhibit it and respond with the correct utterance. This should be accounted for in future studies, where abstract symbols could be used instead of letters in order to avoid multiple phonological representations to be activated.

Notwithstanding, the fact that the articulation-grip congruency effect was replicated with the consonants agrees with the hypothesis that the effect is not specific to the Finnish language. Hence, this finding is in accordance with suggestions that speech-hand connections are evolutionally old, and not

some learned abstract associations, and that their motor networks are partially overlapping [5, 6, 8, 13]. In the field of language acquisition, recent results demonstrate that manual actions can drive tongue protrusions of small children [14]. Comparing these results to our results, which indicated a strong connection between specific grasp actions and articulation, lead us to hypothesize that manual actions could be used to assist speech development in children. Perhaps executing specific grasping actions could guide a child towards proper pronunciation of a certain speech unit.

The results of the current study partially support the hypothesis that the articulation-grip congruency effect is not specific to Finnish, but can be observed in English as well. The effect was replicated with consonants. However, this was not the case with vowels. This could be due to the difference in the correspondence between orthography and pronunciation in Finnish, where it is strong, and English, where it is weak. The current results warrant further studies with English speakers and offer valuable theoretical considerations about the nature of the effect.

## 5. Acknowledgements

We would like to thank Aleksi Rantala for his help analyzing the vocal data. This work was supported by the Academy of Finland, grant 1265610. The work of JL was supported by RVO 68081740/Strategy AV21.

## 6. References

- [1] M. Gentilucci, "Grasp observation influences speech production," *Eur J Neurosci*, vol. 17, pp. 179–184, 2003.
- [2] M. Gentilucci and G. C. Campione, "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*, vol. 6, e19793, 2011.
- [3] V. S. Ramachandran and E. M. Hubbard, "Synaesthesia--a window into perception, thought and language," *Journal of Consciousness Studies*, vol. 8, pp. 3–34, 2001.
- [4] L. Vainio, M. Schulman, K. Tiippana, and M. Vainio, "Effect of Syllable Articulation on Precision and Power Grip Performance," *PLoS one*, vol. 8, e53061, 2013.
- [5] G. Rizzolatti and M. A. Arbib, "Language within our grasp," *Trends in Neurosciences*, vol. 21, pp. 188–194, 1998.
- [6] M. A. Arbib, "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics," *Behavioral and brain sciences*, vol. 28, pp. 105–124, 2005.
- [7] M. Tiainen, L. Vainio, K. Tiippana, N. Komeilipoor, and M. Vainio, "Action planning and congruency effect between articulation and grasping," in *INTERSPEECH 2015 – 16<sup>th</sup> Annual Conference of the International Speech Communication Association*, 2015.
- [8] L. Vainio, M. Tiainen, K. Tiippana, and M. Vainio, "Shared processing of planning articulatory gestures and grasping," *Experimental Brain Research*, vol. 232, pp. 2359–2368, 2014.
- [9] M. Coltheart, K. Rastle, C. Perry, R. Langdon, and J. Ziegler, "DRC: a dual route cascaded model of visual word recognition and reading aloud," *Psychological review*, vol. 108, pp. 204–256, 2001.
- [10] A. M. Liberman, and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, vol. 21, pp. 1–36, 1985.
- [11] L. Vainio, M. Tiainen, K. Tiippana, N. Komeilipoor, and M. Vainio, "Interaction in planning movement direction for articulatory gestures and manual actions," *Experimental brain research*, vol. 233, pp. 2951–2959, 2015.

- [12] B. Hommel, "SR compatibility effects without response uncertainty," *The Quarterly Journal of Experimental Psychology: Section A*, vol. 49, pp. 546–571, 1996.
- [13] G. Maurizio, and M. C. Corballis, "From manual gesture to speech: a gradual transition," *Neuroscience & Biobehavioral Reviews*, vol. 30, pp. 949–960, 2006.
- [14] G. S. Forrester and A. Rodriguez, "Slip of the tongue: Implications for evolution and language development," *Cognition*, vol. 141, pp. 103–111, 2015.