Action planning and congruency effect between articulation and grasping

Mikko Tiainen1, Lari Vainio1, Kaisa Tiippana1, Naeem Komeilipoor1, Martti Vainio1
1 Institute of Behavioural Sciences, University of Helsinki
mikko.o.tiainen@helsinki.fi, lari.vainio@helsinki.fi, kaisa.tiippana@helsinki.fi, naeem.komeilipoor@helsinki.fi, martti.vainio@helsinki.fi

Abstract
Some theories concerning speech mechanisms assume that overlapping representations are involved in programming certain articulatory gestures and hand actions. In previous studies we have shown a compatibility effect between pronouncing or hearing meaningless syllables like [ka] and [ti] and simultaneously performing a power or a precision grip, respectively. The present study investigated whether action selection was necessary for the effect to manifest. The participants were visually presented with a cue for the upcoming manual response. After that, a written syllable “ka” or “ti” was presented at which point the cued grip was performed and the syllable pronounced. There was also a condition, where the grip was cued but only the vocal response was performed. Manual and vocal reaction times were relatively faster when the grip and syllable were compatible (e.g. power & [ka]) rather than incompatible (e.g. precision & [ka]). When no grip was performed (only cued), the effect was still apparent in vocal reaction times. These results suggest that preparation of a manual action is sufficient to influence vocalizations, and also that action selection is not, however, mandatory for this kind of syllable-grip correspondence.

Index Terms: speech production, manual actions, language evolution

1. Introduction
Gestural communication can be considered an integral part of our daily discourses [1], we use gestures every day in conjunction with speech. Darwin [2] already noted that people tend to mimic performed hand actions with their mouth. Similar behavior has been reported with the great apes, so that precise hand movements are often accompanied by mouth movements [3]. From monkey single cell recordings it is actually known that there are cells which are active both when an action is performed with the mouth and when it’s performed with the hand [4]. This kind of a connection can be considered to be evolutionarily old, and several authors have proposed that the foundation for spoken language evolved from a gestural communication system (e.g. [5-7]). Ramchandran & Hubbard [8] have proposed a similar idea by suggesting that some articulations are actually ‘synkinetic’ mimics of hand actions. For example, words that denote smallness would involve narrowing of the vocal tract, such as “little” or “teeny”, and could be synkinetic mimics of precision grip, which is used to grasp tiny objects. Several studies by Gentilucci et al. [9-12] have indeed shown that observing and performing manual actions can influence mouth movements and vocalizations, and vice versa. For example, when watching a larger object being grasped while pronouncing a syllable, the mouth aperture is larger than when the grasped object is small [9]. Likewise, finger aperture is larger when the mouth is open or when pronouncing an open vowel than when the mouth is closed or when pronouncing a closed vowel [12].

Recently, our group discovered a novel connection between hand and mouth actions [13]. Articulations that were produced simultaneously with a grasping action affected the reaction times of those actions. Importantly, the study showed the effects of articulation on grasping to be syllable- and grip-specific. For example, performing a power grip was faster when it coincided with the pronunciation of syllable [ka] rather than [ti]. The opposite happened when participants performed a precision grip, i.e. when pronouncing [ti] the precision grip was performed more quickly than when pronouncing [ka]. Both [ka] and [ti] are meaningless and not related to any words signifying power or precision grip in Finnish.

We proposed that these associations demonstrate that certain articulatory gestures are programmed in a motor network, which partially overlaps with grasp motor representations. For the 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]. For the vowels, the mouth aperture is larger for [a] (an open vowel) than for [i] (a closed vowel). We suggested that articulations involving the tongue body and larger mouth aperture could be thought of 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 analogous of hand movements using the tips of the fingers with smaller hand apertures, like a precision grip [14].

In the present study, we aimed to explore the role of manual action planning in the syllable-grip correspondence effect. It has been reported that observations of large objects are associated with faster power grip responses and viewing smaller objects with faster precision grip responses, but when the response is predetermined and thus no response selection is needed, the effect disappears [15]. Similar results have been obtained in other stimulus-response compatibility studies (e.g. [16]). These results have been interpreted as proof that these effects operate at action planning stage, and if response selection has already taken place prior to stimulus presentation, the stimulus cannot affect the response or this effect is overridden [15]. However, another view is based on the ideomotor compatibility between the stimuli and response [17]. Ideomotor compatibility refers to the similarity between the sensory features of the stimulus and the sensory features of the expected action [18]. According to this view, one of the reasons why prior response selection would diminish the effects is due to the lack of ideomotor compatibility in the task. Brass et al. [17] tested this by using videos of finger
movement as stimuli and asked participants to always respond with the same finger movement when they detected movement in the stimuli. Their results showed that although no response selection was necessary, reaction times were faster when the movement in the video was congruent with the performed response movement [17].

Above mentioned views concerning mechanisms of action planning are linked to perceptually driven action planning processes. However, our view is that similar processes might be responsible for the syllable-grip congruency effect with the exception that it is based on shared action preparation processes between two separate effectors rather than on processing of stimulus and action features in shared sensorimotor representations. To test whether the articulation-grip congruency effect is based on ideomotor or action-selection processes, we employed a paradigm where the manual response was cued at the beginning of each trial. Then as the syllable appeared on the screen, the participants performed the cued grip pronounced the syllable. In this way, the need for response selection was removed for the manual action. If a compatibility effect between syllable and grip is found when no response selection is needed, in the light of the ideomotor compatibility view, it could suggest that there is high ideomotor compatibility between the manual and vocal actions, since it has been argued that in a situation where the task’s ideomotor compatibility is low and response selection requirements are minimal, there is no compatibility effect. Another theory, action-concept model, which is very close to the ideomotor one, holds that stimulus and response codes are part of the same representational domain, and because of this the absence of response selection does not prevent the irrelevant stimulus from influencing the response [19]. This is something that we have already argued could underlie the syllable-grip congruency with an exception to the earlier mentioned common coding of stimulus and action features, the syllable-grip congruency effect might be based on common coding of actions of two separate effectors [14]. Ideomotor compatibility can actually be thought of as motor mimicry which, as already mentioned, has been proposed to be an underlying mechanism for spoken language in general [e.g. 5-8].

Other critical manipulation in our set up was that manual responses were performed according to a go/no-go signal. We used letter case as the go/no-go signal, so for example if the syllable was written in capital letters, the participant performed the manual grip and if the syllable was in lowercase, s/he withheld the manual response (the go/no-go signal mapping was balanced between subjects). The syllable was articulated in go and no-go trials. This experimental setting offered us with the possibility to study the role of action planning from another perspective as well: whether vocal response is influenced by congruency of planned grip action even when the grip is withheld at the time of vocal response. If we would observe the effect even in those conditions, it would suggest that manual execution processes (i.e., control processes) are not required for observing the syllable-grip congruency effect.

2. Methods

2.1. Participants

23 volunteer Finnish-speaking participants (5 male, one left-handed) participated in the study. The participants were aged between 18 to 29 years (mean age 24.6 years). All participants 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 participation. 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 21” LCD computer screen (resolution 1280 x 1080) wearing a head mounted microphone and holding the response devices in their right hand (demonstrated in Figure 1). The two response devices were both 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 12 cm long, 3 cm diameter. As the switches were depressed on each device, there was noticeable tactile feedback. 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 was performed in one block. On each trial a blank screen was first displayed for 2000 ms at the beginning of each trial. A color cue followed the blank screen that was either a green or blue circle. The cue was displayed for 400 ms, followed by a 200 ms blank screen and then the syllable “KA” or “TI” written in black, written in either capital or lowercase letters. The response mapping was such that 13 participants responded with the grip devices only to syllables written in capital letters and 10 participants only to syllables written in lowercase letters. In the go trials, when the syllable appeared on the screen, participants responded as quickly as
possible with the grip device matching the color of the cue and simultaneously pronounced it. Color pairing was also balanced so that 13 participants responded to green cues with a precision grip and 10 participants responded to blue stimuli with the precision grip device. On no-go trials they only pronounced the syllable when it appeared, refraining from any manual response. The stimuli remained in view for 2000 ms or until a manual response was made. Erroneous manual responses were followed by a short “beep” tone. Each stimulus was presented 30 times, which resulted in 240 trials in total (30 x 2 grips x 2 colors x 2 letter sizes). The trials were presented in random order. All participants were given time to practice before the actual experiment, and the experiment was not started until the participant was performing accurately.

2.3. Data and statistical analysis

Vocal data was analyzed using Praat v. 5.3.49. Onsets and offsets were first located individually for each trial as the first observable peak in the acoustic signal for the consonant burst. 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 was subjected to a repeated measures ANOVA with the within factors of syllable ([kɑ] and [ti]) and grip (precision and power) for both manual and vocal reaction times. Post hoc comparisons were performed by means of t-tests applying a Bonferroni correction. A partial-eta-squared statistic employed as effect size estimate.

3. Results

In the go trials for the manual reaction times there was a significant interaction between syllable and grip (see Fig. 2), F(1,21)=16.59, p=.001, ηp²=0.44. Pairwise analysis showed that precision grip was faster when the syllable was [ti] than when it was [kɑ] (463 vs. 532 ms, respectively, p<.001) and power grip was faster when the syllable was [kɑ] than [ti] (473 vs. 524 ms, p=.005). For the vocal reaction times an interaction between syllable and grip was found, F(1,21)=14.72, p=.001, ηp²=0.41 (Fig. 2). [kɑ] was pronounced faster when performing a power grip than a precision grip (530 vs. 587 ms, p=.001) and [ti] was pronounced faster when performing a precision grip than a power grip (530 vs. 578 ms, p=.001).

For the vocal reaction times in the go-no trials we found a similar interaction of syllable and grip as in the go trials, F(1,21)=8.29, p=.009, ηp²=0.28 (Fig. 2). Like in the go trials, when the participants were cued to react with a power grip, [kɑ] was pronounced faster than when the participants were cued to react with a precision grip (510 vs. 525 ms, p=.009). [ti] was pronounced faster when cued to perform a precision grip rather than a power grip, although this difference was only approaching significance (511 vs. 518 ms, p=.084). It should be remembered that although there was always a cue for one of the grips no manual responses were actually made on these trials.

Error rate was small overall, in the go trials 3.00 % for the manual response and 0.58 % for the vocal responses. On the no-go trials error rate was 0.73 %. Due to the very low vocal error rates, only manual errors (go trials) were analyzed. Results of the analysis reflected those of the reaction times. Interaction of syllable and grip was significant F(1,21)=7.17, p=.014, ηp²=0.25. When the required grip was a precision grip, more errors were made when the syllable was [kɑ] than when it was [ti] (4.03 vs. 0.80 %, p=.029). When the required grip was a power grip, more errors were made if the paired syllable was [ti] compared to [kɑ] (5.40 vs. 1.73 %, p=.02). There was also a marginally significant main effect of grip, F(1,21)=4.29, p=.051, ηp²=0.17. More errors were made when the required grip was a power grip rather than a precision grip (3.56 vs. 2.41 %).

4. Discussion

We found the syllable-grip correspondence effect in all reaction time measures: in both manual and vocal reaction times in the go trials, where a manual was performed and in the vocal reaction times in the no-go trials, where the manual action was only prepared. [kɑ] was associated with faster power grip responses than [ti] and [ti] with faster precision grip responses than [kɑ]. Vocally, [kɑ] was pronounced faster when the associated grip was a power grip than when it was precision grip and vice versa for [ti], regardless of whether the grip was actually performed. This means that the effect was observed both in manual reactions when the grip was known beforehand and in vocal reaction times when the grip was only prepared but not executed. The former is in contrast to results of other studies where the participants had prior knowledge of the forthcoming response. Our results thus support the view that activation in articulation related motor representations can influence grip performance even when a grip to be performed has been chosen beforehand. According to the ideomotor theory this might indicate a strong ideomotor compatibility between hand and mouth gestures [17]. From the action-concept models perspective this kind of a result could easily be explained by the shared representational system, as in a shared system it doesn’t matter at which point the activation of the other motor modality occurs, the activation will take place in the shared system thus influencing both modalities [19]. Unlike the ideomotor compatibility, the action-concept model doesn’t stress the similarity between stimulus and response, however it doesn’t exclude it either. Using the ideomotor framework it is quite easy to fit these results to the hypotheses of Ramachandran & Hubbard [8]. If articulations are synkinetic mimes of hand gestures then there would by default be high ideomotor compatibility.

The fact that the effect was found in vocal reaction times even in the case when the manual grip was only prepared beforehand implies that mere grip preparation can influence articulation. This also fits the ideomotor theory, as preparing a grip should automatically activate ideomotorically compatible representations, i.e. in this task a compatible articulation. It would appear that this activation is sustained even if the actual grip is not carried out, possibly offering even more evidence of a high ideomotor compatibility between grip and articulation, where truly just the “idea” (i.e. preparation) of the grip can influence articulations. This result is similar to our previous study, where just hearing or reading syllables [kɑ] and [ti] influenced power and precision grip reaction times [14]. When hearing/reading [kɑ], power grip reaction times were faster than hearing/reading [ti] and vice versa for precision grip reactions. These results completely agree with the current results.

These results support our previous suggestion that representations of grip types and articulatory gestures partially overlap [14]. When the representation of a grip is activated by preparing the grip beforehand, it leads to the activation of the
mechanisms that are responsible for compatible articulation gestures. This activation can then be observed as a motor priming effect in reaction times when articulating those compatible syllables.

5. Conclusions

The results of this study support the view that hand and mouth actions share a common network, where specific activation in one motor property causes specific activations in the other.

6. Acknowledgements

This work was supported by the Academy of Finland, grant 1265610.

7. References